

What hopes for GM food?

The industry based on genetic modification of plants has suffered new setbacks. Europe's public is at the least sceptical, but still potentially accepting, of GM crops, provided progress in technology, regulation and communication is maintained.

It has been a grim year for consumer confidence, not least in genetically modified (GM) foods. The discovery in the United States this week that millions of taco shells need to be recalled owing to contamination by potentially allergenic Aventis corn follows on the heels of the episode in May when the Canadian exporter Advanta Seeds accidentally sent GM oilseed-rape seeds to Sweden and the United Kingdom. Couple that to the fact that the international media-tailored campaigns of crop destruction by Greenpeace and others find ready sympathy with much of the public, and GM proponents might contemplate throwing in the towel. The bizarre but unanimous verdict by a British jury last week that such destruction can be lawful, reportedly influenced by a visible sympathy of the jury with Peter Melchett and his Greenpeace co-defendants, sets the seal on such a pessimistic perspective (see page 438). Is there hope for those who, like *Nature*, adopt an open-minded attitude to the scientific assessment of this technology but also anticipate many benefits from its judicious application?

Uncertainties about the cause of the Advanta contamination are delaying the outcome of a review by the UK Ministry of Agriculture, Fisheries and Food of the crop-separation distances required to achieve acceptably low levels of cross-pollination. But there is no indication that the tens to hundreds of metres recommended for the government's farm-scale trials reflect an underestimation of cross-pollination levels. These trials, the targets of Greenpeace's action, are intended to test the hypothesis that the introduction of maize, sugar beet and oilseed rape, modified for herbicide resistance, will not significantly affect biodiversity. Although it is arguable that they are an excuse for the government to delay the commercial introduction of GM crops, there is no doubt that they could play a leading role in addressing relevant ecological questions. Most positively, they may point to ways in which GM crops could encourage biodiversity.

But for anti-GM fundamentalists such as some of those in Greenpeace, and for many with vested interests in organic farming, such

knowledge is irrelevant. Tying their beliefs to misleading sound bites about potential risks, constantly exploiting fears and misunderstandings about DNA in food, and in the absence as yet of clear benefits from the technology, they have successfully captured much public sympathy.

But that sympathy can quickly evaporate, especially when the public recognizes the manipulation of information — by industry or anti-technology campaigners — for what it is, as has happened in consensus conferences. More potential benefits of GM crops can be expected to emerge, and one can reasonably expect that problems revealed by the science will, as with any technology, lead to appropriate regulation. In short, public confidence can grow, given a chance.

In the meantime, far better public presentation of the state of the science and stricter regulatory precautions are required in Europe. In Britain, the centre of so much debate, the transparency of advice, contrary to general belief, is good (see, for example, <http://www.environment.detr.gov.uk>), as is the willingness of scientists to talk at public meetings. But much of that is ignored in the midst of media heat. The fledgling Agriculture and Environment Biotechnology Commission should try to ensure that, as happened during the BSE crisis, advisory bodies respond rapidly to media debates with informed comment and information.

The thresholds of acceptability of some presence of GM product in organic produce need to be pursued as a priority. It is here that the technical and social issues underlying the inherent conflict between organic farming and other types of agriculture can be resolved with a compromise acceptable to most. This will in turn introduce an additional element of rationality when considering the real and apparent risks posed by GM crops.

And Europe's industry needs to speak rather than, as happens too often, remain silent — and to be less careless than their US and Canadian colleagues in controlling the standards of their products. ■

Triumph, but self-criticism too

The release of an alleged spy at the Los Alamos laboratory comes none too soon, but prompts other questions.

The plea bargain earlier this month that freed nuclear scientist and alleged Los Alamos spy Wen Ho Lee induced jubilation among US scientific leaders. But last week's new disclosure that Lee made copies of the purloined computer tapes of nuclear data makes the accolades premature.

The US government first learned of the additional tapes during the final negotiations for Lee's plea bargain, in which 58 charges were dropped in favour of a guilty plea to a single count of mishandling secret documents. Lee has reportedly insisted they were destroyed. His attorneys only disclosed the copying of the tapes to surprised prosecutors when the plea bargain had been nearly struck, to ensure compliance with plea-bargain stipulations. But it was only uncovered publicly by *Newsweek* last week, setting off a new round of questions in the already troubling nuclear-secret débâcle.

The Los Alamos secrets case appears to engender hysteria like a

nuclear chain reaction: in Congress first, stimulating federal prosecutors to react wildly, providing inaccurate testimony in the effort to convict Lee. And Lee's inhumane treatment during his nine months of incarceration could also be chalked up to that climate. Lee's release has also triggered harsh assessments of government witnesses.

The history of his treatment makes enthusiasm over Lee's release seem entirely justified. Weapons laboratory scientists have much to be proud of, not only for their science (see page 447), but also for so successfully maintaining a balance of scientific openness side by side with the fulfilment of tasks central to the United States' most sensitive defence interests. Yet it is clear from the Lee case that nuclear secrets weren't handled correctly.

The case shows, not for the first time, how scientists need to be ever more scrupulous in adhering to the rules. That is by far the best defence against misbegotten hysteria. ■





AIDS drugs
Should the UK be starting treatment sooner than it does?
p434



Metamorphosis
NSF aims to turn biologists into number crunchers
p435



Star performers
European radio telescopes praised to the skies
p437



Election pledge
Bush vows to stay on course to double the NIH's budget
p438

Europe urged to set up advisory body on research infrastructure

Quirin Schiermeier, Munich

A coordinated approach to scientific infrastructure across Europe came a step closer last week, when a high-level meeting in Strasbourg endorsed a proposal to set up a permanent body to advise the European Commission on such a strategy.

The body would be made up of science administrators from member states of the European Union (EU) and representatives of the commission. It had been proposed by the steering committee of a conference held in Strasbourg on the funding and maintenance of European research infrastructures.

Keen to speed up the planning and implementation of joint facilities, several speakers at the meeting also suggested that the European Science Foundation (ESF) should assess and advise on ways of matching infrastructure needs and resources in the EU.

Such needs include electronic communi-



All aboard: research vessels are one area where a coordinated European policy is needed.

cation networks, databases for the natural and social sciences, telescopes, synchrotron radiation machines, research vessels, and centres in animal and plant physiology.

At present, setting up multinational research facilities under the auspices of the EU requires multilateral — and often cumbersome — agreements between all member states. Many European scientists and science organizations complain that this delays funding decisions relative to the United States and Japan. Science administrators, particularly in the smaller European countries, also complain that it is difficult to get their voice heard in planning new facilities.

"We keep hawking about our plans for a new medium-scale European neutron source," Raoul Kneucker, director-general of the Austrian science ministry, told the meeting. "But obviously there is no relevant panel in the EU where we could go to."

Speakers in Strasbourg suggested that a permanent strategic body could fill the hole. Its role, they said, would be to pave the way for scientific needs to become political funding decisions, in collaboration with the commission and EU member states.

"This would give stability and continuity to the process [and] facilitate a network of decision-makers and funders and hopefully increase the speed of the decision making process," the meeting's steering committee concluded in a formal proposal.

The Strasbourg meeting was organized by the European Commission and the ESF. Senior ESF officials helped to draft the recommendations that were endorsed, including the proposal to set up an advisory group.

Canadian science told to look north

David Spurgeon, Montreal

Canada should rebuild its flagging research programme into northern regions to meet the challenges posed by climate change, rapid population growth, pollution, and social, health and education issues, according to a national task force.

The task force was set up by the country's Natural Sciences and Engineering Research Council (NSERC) and the Social Sciences and Humanities Research Council. Its findings were released last week at the American Association for the Advancement of Science's 51st Arctic Science Conference in Whitehorse, Yukon.

"We have agreed to explore ways of funding parts of it from our existing budgets," said Tom Brzustowski, NSERC's president. "But full implementation will require substantial new funding from the federal government."

In recent years Canada has withdrawn from northern research because of a shortage of funds. "We no longer have the

effective research presence in the north that we need to help safeguard this unique and sensitive environment," says the task force's chairman, Tom Hutchinson of Trent University, Peterborough, Ontario.

The task force had 17 members from university, government and northern communities whose expertise covered natural and social sciences and engineering.

Its report proposes a programme that would establish 14 university chairs in northern research, create 40 northern graduate scholarships and 40 postdoctoral fellowships, support 70 strategic research projects, build partnerships between communities and researchers, and provide money for equipment and infrastructure.

"These measures will allow us to interest young researchers in the north, make sound policy decisions on northern issues, meet major international commitments in the circumpolar region and reassert Canadian sovereignty in the north," says Hutchinson. ■

► http://www.nserc.ca/media_e.htm

Philippe Busquin, the research commissioner, backs the idea of such a group. "A network of pan-European research infrastructures can become a major component of the European Research Area," Busquin told *Nature* (see *Nature* 405, 873; 2000). "I am happy about the high level of agreement reached at this important meeting."

Busquin says that he also welcomes the steering group's recommendation that the commission increase its financial support for building and running common research facilities and giving scientists access to them.



Busquin: backs proposed panel.

This would be good news for two European life-science facilities, the European Mouse Mutant Archive (EMMA) near Rome and the European Bioinformatics Institute (EBI) near Cambridge. Both are in financial difficulties because the rules of the EU's current Fifth Framework Programme exclude such costs from funding.

"The Sixth Framework Programme [FP6], starting in 2003, will certainly contain elements which will help solve the problems of facilities such as EMMA and EBI," predicts Busquin.

Indeed, a proposal to increase infrastructure funding reflects the commission's intention to revise its whole approach under FP6. According to one official, the commission proposes to give less money to small, scattered projects which are often of questionable effectiveness (see *Nature* 404, 695; 2000). The money freed up could then be spent on infrastructure, he says.

"But the commission's role in the funding of infrastructure can only be that of a minority partner, who can hopefully catalyse initiatives," says Busquin, adding that the commission has neither the money nor the power to set up large research facilities without support from member states.

Speaking at last week's meeting, government representatives welcomed proposals to make greater use of the 'variable geometry' approach, under which countries can choose whether to participate in funding particular research facilities.

But an official from the German research ministry points out that governments' enthusiasm for collaboration on research infrastructure is likely to vanish if the costs of new facilities exceed the benefits to the national science base. ■

► http://www.cordis.lu/improving/src/ari_conf_res_infr.htm



The drugs do work: research suggests that early medication helps the immune system fight HIV.

UK faces rethink on AIDS treatment

Karen Birmingham, London

Britain might have to reassess its policy for treating AIDS in the light of new results published this week.

Research at Massachusetts General Hospital and Harvard Medical School in Boston suggests that starting aggressive drug therapy as soon as HIV infection is confirmed is the most effective way to combat the virus. This strategy, say the researchers, allows the patient to develop specialized immune cells that attack the virus (see page 523).

Unlike the United States and most of Europe, Britain delays such treatment until the number of CD4 cells — white blood cells involved in the immune response — drops to below 350 cells per cubic millimetre. This can take several years after initial infection.

The study looked at patients who began taking a cocktail of drugs known as highly active antiretroviral therapy (HAART) within 180 days of infection with HIV. The treatment increased levels of specialized CD4 cells that can destroy HIV-infected cells, and the patients were able to stop taking the medication for up to a year and still suppress HIV levels in their blood. Patients who did not start on HAART until their infection had become 'chronic' showed no signs of this effect.

The case for early treatment is now compelling, says Bruce Walker, director of the Partners AIDS Research Center at Massachusetts General Hospital and one of the paper's co-authors. He is calling for tests that can detect the acute stages of HIV infection to be licensed by the US Food and Drug Administration. This would allow diagnosis roughly two weeks earlier than is possible using current tests.

But Walker and his colleagues stress that their results do not mean that patients can

stop taking their medication without supervision. "It is the drugs that help the immune system to generate the correct CD4 cells," Walker says. "More fine tuning is needed to know how long you need to treat people before you [can] interrupt therapy, and how long that interruption [can] last."

Mike Saag, an AIDS researcher at the University of Alabama at Birmingham, says that the results establish 'proof of principle' that HIV-specific CD4 cells are critical to the body's ability to fight the disease.

"This CD4 cell help is lost very early in the course of infection — within 6 to 12 weeks — and the implication is that early use of antiretrovirals helps preserve this function," he says. "Once preserved, it seems to perpetuate itself even after antiretroviral therapy is stopped for up to a year."

The success of HAART drugs has led to a dramatic fall in AIDS mortality in the western world, but Britain has resisted giving otherwise healthy people these powerful drugs. Over the past two years, as toxic side effects of the long-term use of HAART have surfaced, such caution has seemed justified. Patients can suffer disfiguring body-fat redistribution that might be associated with diabetes and cardiovascular diseases.

Although the new data suggest that immediate treatment could reduce the time that patients spend on the drugs, Jonathan Weber, professor of communicable diseases at Imperial College School of Medicine in London, and a proponent of delaying therapy, urges caution.

"HIV is a long illness, and just because someone looks great at two years, it doesn't tell you what will happen 14 years later," he says. But he admits that "the time is right to start a randomized clinical trial of when it is best to start treatment." ■

NSF aims to inject more maths into biology

Rex Dalton, San Diego

In a bid to bring more direction to a somewhat chaotic field, the US National Science Foundation (NSF) has launched an ambitious plan to increase biologists' use of mathematical and statistical techniques.

The agency wants researchers to build these techniques into their grant proposals. The scheme's initial budget will be at least \$3 million, say officials, and future budget requests could boost agency-wide support to \$100 million.

It is hoped that the initiative will provide tools for analysing the large data sets that will come from research programmes such as the planned National Ecological Observatory Network, a system of long-term monitoring facilities. The eventual goal is to find definitive answers to questions on population dynamics, dispersal of species over large areas, and the impact of ecological variables.

NSF officials say the fund will initially target neurobiology, animal behaviour, ecology and evolutionary physiology. New numerical techniques could be applied at every level from genes to ecosystems.

The NSF has sent out notices about the initiative, known as Quantitative Environmental and Integrative Biology. The agency's biological sciences directorate and its mathematics and physical-sciences directorate will use their core programmes to encourage the funding of research incorporating the new approach.

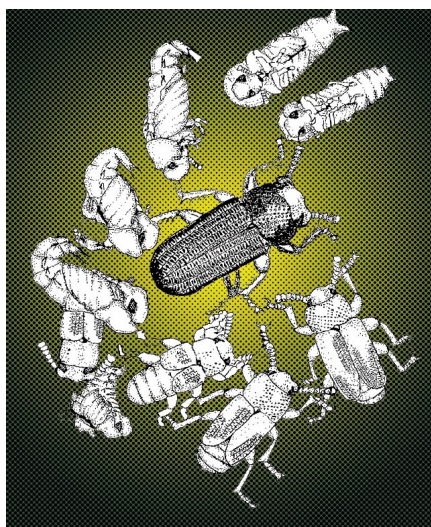
The notice says that the time is ripe to use existing mathematical tools and develop new approaches to speed up progress towards understanding and predicting environmental and integrative biology.

Principal investigators are being encouraged to train undergraduates and graduates in the new techniques, with a view to them becoming involved in the research. NSF officials see this as a way to create a much-needed generation of scientists skilled in both biology and mathematics.

Earlier this month, at an NSF-sponsored workshop at the University of California at San Diego's Supercomputer Center, mathematicians, statisticians and biologists came together to give the agency guidance on the types of project it should fund.

According to Margaret Palmer, director of the NSF's division of environmental biology and an aquatic ecologist at the University of Maryland, the consensus was that the initial targets should be population genetics, ecological system restoration, and animal physiology and behaviour.

People were excited about the potential for progress at the interface of environmental biology and mathematics, says Alan Hastings, an ecological mathematician at



Adding up: maths has been used successfully to predict changes in flour-beetle populations.

the University of California at Davis, who organized the workshop. The most interest, Hastings said, surrounded the issues of randomness and variability.

Scientists have had some success at developing models to analyse population data, for example, with the flour beetle, voles and Dungeness crabs, predicting population parameters and working out what determines species populations.

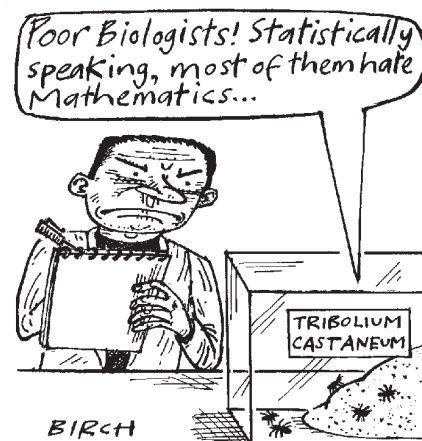
But difficult challenges remain, particularly in studying species spread over large areas. Many serious ecological questions involve space, such as the spread of invading species and genetically modified traits through native populations, says Tony Ives, a University of Wisconsin zoologist who participated in the workshop.

For wide-ranging populations, data at any one location are likely to be sparse, leading to statistical problems with characterizing populations. One attempt to address this involved a team of mathematicians and ecologists who worked with laboratory populations of the flour beetle *Tribolium castaneum*.

This produced a seminal paper (see *Nature* 375, 227–230; 1995) showing how a mathematical model could be used to predict changes in a species where cannibalism leads to wildly differing population growth rates.

At the NSF workshop, members of the team — Brian Dennis of the University of Idaho, Robert Desharnais of California State University at Los Angeles and Shandelle Henson of the College of William and Mary in Williamsburg, Virginia — explained how their techniques can be applied to species in the wild, translating biology into mathematical language.

► <http://caldera.calstatela.edu/nonlin>



France backs IT and life sciences

Declan Butler, Paris

France's spending on civil research and development will place a greater emphasis on information technology and the life sciences next year, science minister Roger-Gérard Schwarzenberg revealed last week.

Schwarzenberg also launched a plan to recruit young scientists which will be used to direct strategic goals towards these two fields and reduce the heavy burden of salaries on research agencies' budgets.

The FF55.8 billion (US\$7.5 billion) budget for 2001 is a 2.2% increase on this year. But overall, science came only seventh in additional spending — the ministry of environment, for example, got an 8% boost.

There will be 265 new full-time posts for researchers and technicians in public

research agencies, compared with 150 in 1999 and just 18 last year. This is partly because half of France's current researchers will retire between 2004 and 2010.

Of the new posts, 116 will go to the National Institute for Research in Computing Science and Control (INRIA). Many of the 70 jobs at the Centre National de la Recherche Scientifique (CNRS) will go to a department of information technology, which is expected to be formally created alongside the seven existing CNRS departments this week.

INRIA's 12.7% rise tops the budget handouts to the national agencies. The national biomedical agency INSERM got a 4.2% increase and CNRS got 0.9%.

Schwarzenberg said that 25% more public researchers would come into



Berger: sees interdisciplinary research as key to the future of French science.

information technology over the next five years. Given that 74 of the new jobs will go to INSERM, and that many of the non-computing posts at the CNRS will go to the life sciences, the harvest for other disciplines has been meagre. Jacques Fossey, secretary-general of the main researchers' trade union, the National Union of Scientific Researchers (SNCS), is among those critical of this imbalance.

But Geneviève Berger, the new director-general of the CNRS, says that she will not neglect physics and chemistry. Rather, in a bid to boost interdisciplinarity, she wants to see more researchers in these areas applying for jobs in life-science research.

Berger is also planning a bioinformatics programme to link the life sciences with the proposed information technology department. And she wants to introduce within six months a new 'matrix' model for research at the CNRS, with projects cutting across disciplinary boundaries.

Berger sees the wave of retirements as a golden opportunity to make the management of research more flexible. At present, over three-quarters of the CNRS budget goes on salaries, leaving little for equipment and supplies. She would like to see this cut to 60%.

Berger does not intend to challenge researchers' status as civil servants. But she wants fewer full-time researchers and more temporary posts, to allow more foreign researchers to work at the CNRS. A system of three-year postdoctoral fellowships, renewable once, is also under consideration.

Having studied the budget figures, the SNCS claims that, despite the government's emphasis on job creation, it intends to increase spending on research infrastructure by cutting the wage bill. Describing this as "scandalous", the union is urging the National Assembly to amend the budget bill when it comes up for approval later in the autumn. ■

Disagreements derail US bill to double research funding...

Colin Macilwain, Washington

The hopes of US scientific societies that Congress would pass a bill to double research spending in each of the major non-military research agencies collapsed last week, when the chairs of the relevant committees in the House and Senate failed to reach agreement on the bill's contents.

After four months of negotiations between the staff of James Sensenbrenner (Republican, Wisconsin), chair of the science committee in the House of Representatives, and Senator Bill Frist (Republican, Tennessee), Sensenbrenner refused to support Frist's proposal in the House — almost certainly ending its prospects of passage this year.

Sensenbrenner wants a bill supporting increased research funding in information technology, whereas Frist backs a much broader bill supporting all types of research.

In a detailed letter sent to Frist on 19 September, Sensenbrenner said that Frist's proposals "would provide little support for scientific research funding while undermining the science committee's ability to operate as an effective legislative entity".

If passed, Frist's bill would not guarantee a doubling of each agency's budget, but it would set an overall target for expenditure over several years. Sponsors hope this would influence the appropriations committees when they allocate the agencies' annual budgets.

Sensenbrenner is thought to oppose the bill mainly on the grounds of cost. But his letter argues that it would also reduce the prospects of Congress considering detailed legislation for science agencies — such as the

NASA reauthorization bill passed this year — and would further reduce his committee's influence on the appropriations process.

Two days after Sensenbrenner's letter, the Senate unanimously passed Frist's bill, the Federal Research Investment Act. But without Sensenbrenner's support, it has little chance of being considered by the House during this session.

In an angry reply to Sensenbrenner, Frist indicated that the science committee chairman had failed to live up to promises made in their negotiations. "You simply are holding the Federal Research Investment Act up to a different standard than you do your own committee bills," he wrote. He added that "anything less" than Sensenbrenner's support for the act "would be in direct contradiction to our previous negotiations".

Scientific societies, including the American Physical Society and the American Chemical Society, have been trying to build support for the measure for over three years (see *Nature* 394, 5; 1998). They believe the bill will ensure research agencies supporting the physical sciences — such as the Department of Energy — will not be left behind as Congress backs the rapid expansion of life sciences at the National Institutes of Health.

Some science lobbyists have said that the bill should be abandoned if it fails in the current session. But according to staff, Frist will pursue it in the new Congress next year. Whatever happens in November's election, Sensenbrenner is not expected to chair the science committee. His departure will remove the largest single obstacle to the bill's passage. ■

...as physical sciences plan lobbying body

Scientific societies and research foundations plan to set up a lobbying body in Washington to boost public support for investment in the physical sciences, mathematics and engineering.

Impressed by the success of Research!America, a society that lobbies for the life sciences, various scientific bodies have provided around \$100,000 with which to prepare a business plan. The interim director of the body is to be Merrilea Mayo, a materials scientist from Pennsylvania State University.

The attempt to establish the

as-yet-unnamed outfit has been spearheaded by Mary Good, a professor at the University of Arkansas and former under-secretary of technology at the commerce department in President Bill Clinton's administration. The physical sciences, says Good, "need a continuous presence" in Washington. "We need to have the community speak with one voice, instead of 30 or 40."

Those involved in the effort say the precise role of the new body has yet to be defined, and that its success will depend on how much support it can attract



Good: believes the physical sciences need a single voice.

from sectors of industry that want the federal government to invest more in the physical sciences. **C.M.**

Antibiotic resistance must be monitored, US Senate is told

Paul Smaglik, Washington

The United States needs a better monitoring system to combat the threat posed by the rising number of bacteria resistant to existing antibiotics, federal officials told a Senate hearing last week.

Apart from drug-resistant tuberculosis, which is tracked in all 50 states, monitoring of antibiotic-resistant microbes remains relatively fragmented.

According to the limited data available, drug resistance can vary widely, even within states. In Maryland, for example, 15% of the strains of *Streptococcus pneumoniae*, the bacterium that causes pneumonia and meningitis, are resistant to penicillin. In five Tennessee counties, resistance is seen in 38% of the strains.

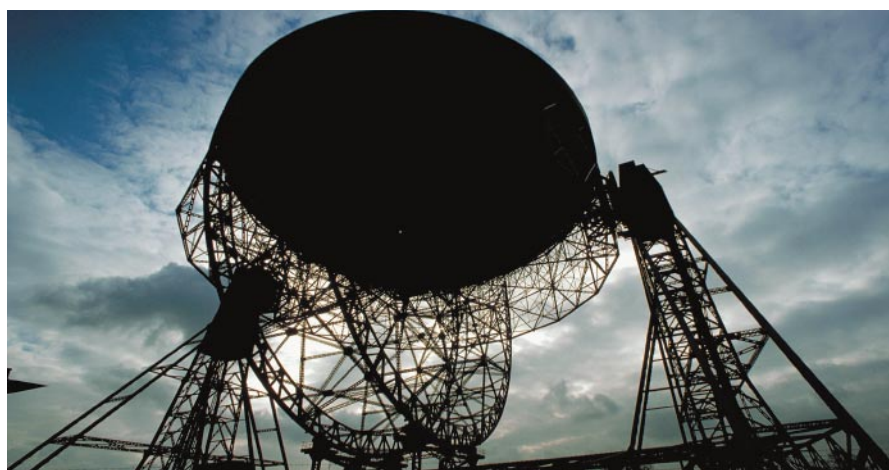
But this kind of reporting is not the norm, Jeffrey Koplan, director of the Centers for Disease Control and Prevention (CDC), told a hearing of the Senate Appropriations Subcommittee on Labor, Health and Human Services, and Education. Better monitoring, he said, as well as more research into how microbes develop resistance, needs to be stressed in the next federal budget.

New forms of drug resistance continue to develop, Koplan explained. For example, a CDC report last week identified two strains of *Neisseria gonorrhoeae*, the bacterium that causes gonorrhoea, that have recently become resistant to some of the drugs used to treat the condition.

Jane Henney, US Food and Drug Administration (FDA) commissioner, agreed that surveillance is a key part of fighting drug-resistant microbes. Spotting resistant strains earlier allows doctors to prescribe different antibiotics, which is important, as using ineffective antibiotics promotes further resistance.

Doctors need to be more aware of the dangers of prescribing antibiotics unnecessarily, and patients should be educated about their proper use, Henney said. She also noted the need for faster development of new vaccines and antibiotics.

It remains unclear which strategies will be given highest priority in the US administration's next budget request. But the FDA, CDC, National Institutes of Health and seven other federal agencies formed a task force on drug resistance last summer. This has since released a series of recommendations, including the need for a better system to monitor drug-resistant pathogens.



Team member: the Jodrell Bank radio telescope, one of 18 telescopes that form part of the network.

Report praises European radio telescope network

Alison Abbott

A European network of radio telescopes that allows small celestial objects to be imaged at very high resolution is a major scientific success, according to a review published last week. But the report also says the network needs secure, long-term funding if it is to provide optimal value to its growing user community.

The European Very-Long-Baseline Interferometry Network (EVN) was set up on an ad hoc basis in the 1980s by 14 astronomy institutes. The network's telescopes dedicate up to 30% of their observation time to simultaneous data-gathering, greatly increasing the resolution of observed images.

The resolving power of a telescope is limited by its physical diameter. Very-long-baseline interferometry is a technique that, by linking up distant telescopes, increases the diameter of their collecting area to the distance between them.

The EVN links 18 radio telescopes in this way, including three of the world's largest, which have diameters of 100 metres (at Effelsberg, near Bonn), 94 metres (Westerbork, the Netherlands) and 76 metres (Jodrell Bank, United Kingdom).

Last week's review was commissioned by the Netherlands national research council on behalf of the research councils involved in the EVN, and was carried out by the European Science Foundation. It was aided by an international panel chaired by Jens Erik Fenstad, a mathematician at the University of Oslo.

The report's praise for the network extends to astronomers working in wavelengths other than radio. One example is the EVN's detection of 50 supernovae in M82, a relatively close galaxy whose central regions are "dusty".

"We can watch the gas that forms when a supernova remnant expands, allowing us to study the physics of star formation and star death," says Harvey Butcher, director of the Netherlands Foundation for Research in Astronomy at Dwingeloo.

Although initial funding for the EVN came from its participating research institutes, the network has since been designated a European Union Large Scale Facility. This allows the European Commission to support scientists' access to the facility. The network has also raised funding from various sources to develop data-processing technologies.

Last year, for example, the EVN brought online the world's largest 'correlator', which merges data the telescopes have collected on magnetic tape and converts them to images. The correlator is based at the Joint Institute for VLBI in Europe (JIVE) in Dwingeloo.

At present, however, half of JIVE's operating budget is paid for by a consortium of seven national research institutes through short-term bilateral agreements which expire at different times. The rest comes from sources that have been gathered by EVN institutes.

Such a funding policy basis is "too frail" for a key facility such as JIVE, says the report. According to Phil Diamond, a director at the Jodrell Bank Observatory and a member of the report's panel of experts, national research councils must at least double their contribution to JIVE — currently around 700,000 euros (US\$615,000) — and operate through a single, multinational agreement.

Further sources of income would also be needed, he says, if EVN-JIVE is to develop its technologies and services.

Butcher believes that the report should help EVN-JIVE in its fight for a simple and stable source of funding.

http://www.esf.org/update/news/00/jive.htm

US taco shells found to contain unlicensed modified corn

Washington Kraft Foods announced the recall of 2.5 million boxes of 'Taco Bell' taco shells from US supermarket shelves last week after an environmental group established that the product contained a genetically modified strain of corn that has not been approved for human consumption.

The corn shells contained a protein from the bacterium *Bacillus thuringiensis*, Cry9C, which does not occur in the human food chain. The Food and Drug Administration had therefore been unable to establish definitively whether the protein causes allergies, and has not approved it.

Corn containing the protein, which is marketed by Aventis under the brand name StarLink, is approved only for animal consumption. There have been no reports of illness resulting from consumption of the recalled taco shells.

UK farmers seek protection after protesters walk free

London Britain's farmers have asked the government for greater protection for field trials of genetically modified (GM) crops after criminal charges were dropped last week against 28 people who had admitted destroying six acres of experimental GM maize at a site in Norfolk in April last year.

Lord Peter Melchett, the executive director of Greenpeace and one of those arrested, said that, with several farmers who had volunteered to participate in the trials now dropping out, doubts were growing over whether the government would be able to gather enough data to make future tests valid.

Bush pledges more big budget rises for NIH



Bush: plans to continue 15% annual increases.

Washington US presidential candidate George W. Bush promised last week that, if elected president in November, he will double the budget of the National Institutes of Health (NIH) from the \$14 billion it was in 1998 to \$27.3 billion in 2003. The NIH budget has been on course for such

doubling over the past three years, and Bush, the Republican presidential nominee, proposes to keep it rising at the same rate — which would imply annual 15% increases — for a further five years.

"I will lead a medical moonshot to reach far beyond what seems possible today and discover new cures for age-old afflictions,"

said Bush in an address in Florida. Vice-president Al Gore, the Democrat presidential nominee, has promised to double the portion of the NIH budget that goes for research grants between 2000 and 2006 (see *Nature* 407, 118; 2000).

US scientists wary of leaving industry for government

Washington Top scientists from industry are being deterred from accepting key positions in the US government, according to the National Academy of Sciences. It says they fear lengthy delays in their confirmation process, and are worried by the growing array of restrictions on their activities during and after office.

In a report published last week, an academy panel chaired by Mary Good, former under-secretary of commerce, called on the new administration to get the appointment process moving ahead of the November election, to make 80–90% of its appointments in the first four months of 2001, to streamline security checks, and to keep appointees better informed on the status of their nominations.

Spain prepares to send bushman home

Barcelona The stuffed body of a nineteenth-century bushman, whose display in a museum near Barcelona provoked international protests, will be sent home to Botswana for burial early next month. The body was moved to Madrid two weeks ago, where experts are preparing it for repatriation.

The body went on display in 1916 at the Darder Natural History Museum in Banyoles. In 1991, a Haitian doctor living in Catalonia denounced the exhibit as racist. The Spanish government and the town agreed in June to return the body to Africa, stating that it had been removed from the museum out of "respect to the African immigrants who live in this town".

Europe is neglecting basic biotech, says study

Munich Although biotechnology enjoys generous public funding everywhere in Europe, its less commercial aspects are neglected, according to a study carried out at the Fraunhofer Institute for Systems and Innovation Research in Karlsruhe, Germany. European countries spend around US\$1.8 billion per year on biotechnology, with medical and pharmaceutical research as the top priorities.

The study predicts two future trends. Larger countries, including Germany, France and the United Kingdom, will increase their investment in basic research, it says, but the rest are likely to concentrate public funding on applied research. Also, it says, Britain is

the only country in which support for environmental applications of biotechnology is likely to increase substantially.

Russia to create prize for industrial science

Moscow Russia's minister of industry, science and technology, Alexander Dondukov, has told the cabinet that his ministry intends to create a prestigious prize for achievement in industrial research. "It should be a Russian 'Oscar' for industrialists and scientists," he said, "and the money a laureate gets must be very impressive."

The first prize is expected to be awarded in December. The minister said that the prize has been made possible by economic growth. "In particular, the metallurgical industry alone now has an annual income of more than US\$1 billion," he said. "Some enterprises in this industry have agreed to sponsor the new nationwide prize."

Europe gives money to map methylation of the genome

Brussels The European Commission is to provide 1.2 million euros (US\$1 million) to the Human Epigenomics Project, an initiative to map the DNA methylation patterns of the entire human genome. DNA methylation — the addition of a methyl group to an individual DNA base — acts as a switch by which specific genes can be turned off. The detection and interpretation of DNA methylation patterns, called epigenomics, is expected to increase understanding of complex diseases.

The project was started in 1999 and includes a consortium of scientists from the Sanger Centre near Cambridge, the Centre National de Génotypage in Paris and the Berlin-based biotechnology company Epigenomics, as well as three additional German research institutes. It will make its data publicly available, and its first step will be a methylation map of the major histocompatibility complex, which is of key importance in immunology.

India plans reusable microgravity satellites

New Delhi The Indian Space Research Organization (ISRO) is proposing to build a series of reusable satellites for conducting microgravity experiments. The satellites are the second phase of the national programme of microgravity research, for which ISRO earlier this month invited proposals from the Indian scientific community.

In the first phase, experiments will be conducted in modules dropped from balloons, launched by the National Balloon Facility in Hyderabad. ISRO says it can build the recoverable satellites in three years once it gets formal government approval.

A springboard to success

Can Brazil build on its achievement of completing the first genome sequence of a plant pathogen? That may depend on its willingness to reform its universities, say Colin Macilwain and Ricardo Bonalume Neto.

When Fernando Perez, scientific director of FAPESP, the Brazilian state of São Paulo's science funding agency, launched an ambitious project to sequence the genome of a bacterium that attacks orange trees, his main objective was to train the state's scientists in the techniques of genomics. But now that the sequence of *Xylella fastidiosa* has been published in *Nature*¹, the project is being hailed as a great deal more than a successful training programme.

The *Xylella* experience proved that Brazil's researchers can hold their own in the big league, turned its leaders into media personalities, and has seen them fêted by the nation's president. "It mobilized public opinion to accept the possibility that we can achieve many things in science," says Carlos Pacheco, deputy minister of science and technology in the Brazilian federal government. The challenge now, for scientists and policy-makers alike, is to maintain the momentum and ensure that other talented Brazilian researchers get a chance to compete at the highest levels.

Encouragingly, the federal government is following São Paulo's lead and stepping up its investment in research. But many scientists warn that, if this spending is to bear fruit, there will have to be major reforms to Brazil's inflexible university system. Indeed, given the rate of progress in genomics, it will be a challenge even to maintain the country's current position in that field. "We have entered a high-speed freeway," says Perez. "We have to adjust to the situation, and make the right choices."

The *Xylella* sequence was undertaken by a consortium called ONSA, the Organization for Nucleotide Sequencing and Analysis, at a cost of US\$13 million. The project's leaders carved out their genomics niche by picking an organism that is important to the local economy, and turned to a collaborative model, linking many small labs.

The use of a consortium is not new — the European yeast-genome project took a similar approach². But the subsequent development of the Internet has made such collaborations easier. And the model proved

especially appropriate in Brazil, where the fierce competition between individual labs that drives North American and European science is not so evident. "We are not competing with one another necessarily. The push comes from competing with scientists abroad, as a group," says Andrew Simpson, an English-born geneticist at the Ludwig Institute for Cancer Research in São Paulo, who helped lead the project. "We may have hit on a solution for how developing countries might develop their science."

Opportunities and obstacles

Brazilian scientists in other disciplines are already trying to think of ways to replicate the ONSA model, no doubt inspired by the outpouring of national pride that greeted the *Xylella* project's appearance on the cover of *Nature* — recognition befitting ONSA's achievement. Simpson and Fernando Reinach of the University of São Paulo, another leader of the project, were suddenly the centre of public attention. "It has had an enormous impact," says Hernan Chaimovitch, vice-president of research at the University of São Paulo. "The attitude of people in the media to science has changed dramatically."

But as the euphoria dies down, serious questions are being raised about the ability of

Cover stars: publication of the genome sequence of the citrus pathogen *Xylella fastidiosa*, which is spread by a leafhopper (above), made celebrities of Brazilian geneticists.

Brazil's university system to support similar, internationally competitive projects. The rumblings surfaced last month with the publication in *Nature*³ of a letter from Tomas Prolla, a Brazilian geneticist working at the University of Wisconsin at Madison. Prolla was responding to suggestions that the success of the *Xylella* project might convince young Brazilian researchers to stay in the country rather than leave for greener pastures abroad⁴. "The rigid bureaucracy of Brazilian public universities turns the simplest transaction into a nightmare," he wrote. "For most Brazilian scientists with academic positions in the United States or Europe, returning home remains akin to academic suicide."

Prolla cites poor academic salaries and a lack of support for young faculty members as particular obstacles. And although other scientists are less outspoken, they echo many of

All for one, one for all: Simpson believes that scientists at different Brazilian institutes are competing as a group with researchers in other countries, rather than with each other.



his criticisms. Reinach argues that the universities are too rigid in their allocation of resources to departments — refusing, for example, to employ computer scientists in biochemistry departments. The genome projects “are putting pressure on the system — which is a good thing,” he says. The *Xylella* project allowed young researchers to bypass the universities’ rigid career structure, Reinach adds, and created a new “culture of accountability” by paying research teams according to their results, rather than seniority. “I hope that doesn’t go away.”

Room for improvement

“The University of São Paulo is organizing itself to respond faster to this kind of challenge,” responds Chaimovitch. Last month, for example, it announced the establishment of a Centre for Bioinformatics, initially with 16 staff, which he expects to serve as the launch pad for a graduate programme in the discipline. Reinach agrees that the strongest universities are making some progress. But overall he laments that the response has been “very, very slow”.

For the universities, lack of resources remains a huge problem, in particular when it comes to retaining technicians and the young researchers on whom the future of Brazilian science depends. With universities unable to supply enough permanent support staff, scientists in São Paulo state often hire technicians on two-year FAPESP ‘fellowships’, which seldom lead to permanent jobs.

“This is not a good situation,” says Jesus Ferro of the State University of São Paulo at Jaboticabal, whose lab is a member of ONSA. Talented young scientists, meanwhile, are lured abroad or take faculty positions at Brazilian universities that involve little research — because this is their best chance of getting a permanent job. “We can attract students,” says Chaimovitch. “The problem is keeping them from going to Europe and the United States and being paid royally.”

The *Xylella* project has also shone a spotlight on the great inequity of Brazilian science: the relative wealth of São Paulo state compared with the rest of the country³. Heavily industrialized São Paulo dominates the Brazilian economy, and the state’s scientists benefit from statutes that give FAPESP 1% of São Paulo’s tax revenues.

But having seen FAPESP’s star rise so dramatically as a result of the *Xylella* project, the federal government is eager to get a piece of the action. Even before the *Xylella* sequence was published, it had announced a plan to double its spending on university research next year, boosting funding by US\$600 million through 11 goal-orientated programmes. This month saw the announcement of the last three of these — in health, environmental and aeronautics research. The federal government is also launching its own genome effort, establishing five



Hungry for change: Reinach laments that only “very, very slow” progress is being made towards a more dynamic university culture.

sequencing networks similar to ONSA.

ONSA, meanwhile, is moving to build on the *Xylella* project. Twenty-one grants — ranging from US\$20,000 to US\$600,000 — have been released to laboratories to investigate the function of the pathogen’s genes. “Functional genomics is another challenge for us, because it is not something that is so easy to do as a collaboration,” says Perez. On the sequencing side, ONSA has decided to concentrate on plant pathogens. Three such sequences are already being worked on, and FAPESP is selecting ten more.

A sequence of sequences

The first of the three ongoing projects is an effort to sequence another strain of *Xylella fastidiosa* that causes disease in vineyards in southern California. The US Department of Agriculture will pay half of the US\$500,000 that the project will cost. “We hope that the genes will open up some additional opportunities for disease-management strategies,” says Edwin Civerolo, head of the crop pathology and genetics unit at the University of California at Davis.

Meanwhile, a team led by Ferro has almost completed the sequence of *Xanthomonas citri*, which causes citrus canker. This species also attacks many other commercially important crops, including rice, beans, cotton and grapes, raising prospects for future international collaborations. The third project, headed by Luis Aranha of the University of São Paulo at Piracicaba, is focusing on *Leifsonia xyli xyli*, which attacks sugar cane. And FAPESP is planning to use

Expressed Sequence Tags (ESTs) — fragments of DNA derived from messenger RNA — to identify genes of scientific and agricultural interest from sugar cane itself. This project will involve researchers outside São Paulo, including the northeastern Brazilian states of Alagoas and Pernambuco.

FAPESP and the Ludwig Institute are also supporting a US\$10 million cancer genome project to identify ESTs associated with certain types of tumour, in particular those of the head and neck, which are common in Brazil. The researchers are using a technique called open reading frame EST sequencing, developed by Simpson’s team, which homes in on sequences that are translated into proteins⁶. The project already accounts for about 20% of cancer-related ESTs deposited in public databases worldwide, and Simpson aims to have produced a million ESTs by the end of the year. Meanwhile, other researchers want to turn ONSA’s sequencing machines loose on parasites that cause disease in Brazil, such as the blood fluke *Schistosoma mansoni*, and *Trypanosoma cruzi*, which causes Chagas’ disease.

Whether Brazilian scientists in other disciplines will follow ONSA’s example remains to be seen. But the consortium’s leaders are confident that the momentum in genomics will not be lost. “A lot of very young guys were on the *Xylella* project, and it turned up some very good people,” says Reinach. “They learned the technique and are now using it for their own different things.” The experience was “absolutely fantastic,” says Sandro de Souza, one of these young scientists, who now runs the bioinformatics programme at the Ludwig Institute in São Paulo. ■

Colin Macilwain is Nature’s Senior US Correspondent; Ricardo Bonalume Neto reports for Nature from São Paulo.

The federal government has announced a plan to double its spending on university research.

1. Simpson, A. J. G. *et al.* *Nature* **406**, 151–157 (2000).
2. Goffeau, A. *et al.* *Nature* **387** (suppl.), 1–105 (1997).
3. Prolo, T. *Nature* **406**, 826 (2000).
4. *Nature* **406**, 109 (2000).
5. Macilwain, C. *Nature* **398** (suppl.), A16–A18 (1999).
6. Dias Neto, E. *et al.* *Proc. Natl Acad. Sci. USA* **97**, 3491–3496 (2000).

Cheap and cheerful circuits

Silicon remains the computer chip industry's material of choice. But for simpler circuitry, it could soon have some surprising rivals. David Voss talks to the scientists who are trying to make electronics go organic.

It is easy to get the impression that research in electronics is all about pushing the limits of information processing. The fanfare that greets the launch of each new, ultrapowerful computer chip certainly makes it seem that way. But another trend is quietly gathering force: a push to exploit the modest end of cost and performance. For the past few years, materials scientists and device engineers have been collaborating to develop a new breed of cheap and cheerful electronic components. Their working medium: not silicon, nor metals — but plastics and smaller organic molecules.

The goal of organic electronics is to make devices that are cheap, flexible and just smart enough to perform the task in hand. Electronic price tags, flexible computer screens and disposable cell phones are among the potential applications. Their components will be created not in billion-dollar fabrication facilities, like today's silicon chips, but using printing technologies little more sophisticated than those found in a desk-top inkjet printer.

"People talk about microelectronics," says materials scientist Francis Garnier of the CNRS Laboratory of Molecular Materials in Thais, part of France's national research agency. "But you can also have macroelectronics — low-end and inexpensive." And in the past year, several groups have announced striking results that have boosted the field's prospects. "It's been a rich haul," says Ananth Dodabalapur, whose team at Lucent Tech-

nologies' Bell Labs in Murray Hill, New Jersey, has been at the forefront of this effort.

Manufacturers love plastics — they are durable, cheap and easy to process. But in electronics, plastics have largely been confined to providing the boxes that house the circuitry and insulating wrapping for the conductors and semiconductors that do the real work. Some polymers, however, can perform as reasonable semiconductors, or even be fully conducting. The polymers involved all have a 'conjugated' backbone, consisting of alternating double and single bonds. These bonds are delocalized, which means that electrons can move along the molecule, much as they do through silicon.

Reassuringly, it seems that the basic physics of organic semiconductivity is much the same as in more familiar semiconductors. "That has given a huge impetus to the field," says Richard Friend, a materials scientist at the University of Cambridge. "We're not just mucking around with dirty materials on the fringe."

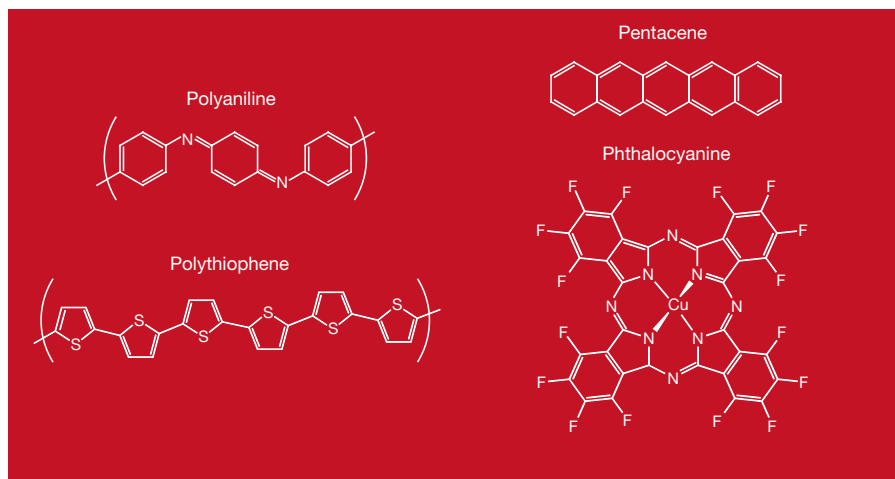
Flexible computer screens and disposable mobile phones are among the potential applications.

An early hint of the possibilities offered by plastic electronics came in 1994, when Garnier's group reported making the first all-polymer transistors¹. Laying down the molecules using simple printing technology, rather than sophisticated techniques such as high-vacuum deposition or photolithography, Garnier and his colleagues fashioned rudimentary transistors out of the organic semiconductor polythiophene. Organics and polymers had been used in transistors previously, but this was the first time transistors — including their electrodes — had been made completely from plastic. As a result, the transistors kept on working even when the substrate was bent through angles of 90°.

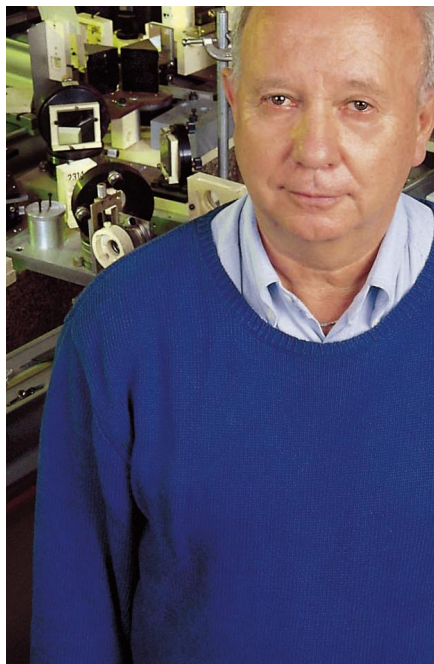
Towards integration

But transistors are not much use by themselves. The next step is integration — linking many transistors together to process information. Dago de Leeuw and his colleagues at Philips Research Laboratories in Eindhoven, the Netherlands, took the first big step forward in 1998, producing a plastic integrated circuit in which batches of transistors were hooked together into a functional unit². After making individual transistors, and then simple logic gates, the researchers eventually fashioned a 15-bit code generator containing 326 all-polymer transistors. This is a device that creates a unique digital code for use as an identifying number, rather like a barcode. And it has already found a commercial niche: Philips has incorporated the circuit into identification tags that are used as anti-theft stickers for retail goods.

De Leeuw and his colleagues achieved their results using a kind of photochemical patterning. First, they coated a substrate with a mixture of polyaniline and a photochemically reactive chemical, spinning the substrate rapidly so that the coating materials ended up as a film less than one micrometre thick. The photoreactive compound had the handy property of turning the normally conductive polyaniline into an insulator when exposed to ultraviolet radiation. So when the substrate was exposed to ultraviolet light through a mask, the irradiated parts of the film underwent a chemical transformation that boosted electrical resistance by 11 orders of magnitude. With the basic geometry of the circuit laid down in



Conjugal bliss: polymers (left) and smaller organic molecules (right) that can act as conductors or semiconductors have a 'conjugated' structure of alternating single and double bonds.



Garnier: trying to make cheap "macroelectronics".

this way, the researchers deposited a film of their organic semiconductor, polythienylenevinylene, followed by an insulating layer of polyvinylphenol. Finally, the researchers punched tiny holes through the substrate, into which polyaniline seeped, connecting the output of one transistor within the circuit to the input of the next.

In February this year, Dodabalapur and his colleagues at Lucent raised the bar³. Their circuits are not made entirely of polymers — the component transistors are connected using nickel and gold. But they use much less power than the Philips device, thanks to a technique called complementary logic, borrowed from conventional silicon microelectronics. Semiconductors come in two flavours, n- and p-type, depending on whether they have a surplus or a deficit of electrons. Transistors can be made out of either, but device engineers discovered long ago that circuits made only of n- or p-type transistors draw lots of current. Logic elements made from a combination of the two, on the other hand, draw very little except when they are switched on or off.

From science to gadgets

Taking a complementary-logic approach, the Lucent team created circuits containing 864 organic transistors. They used two different organic semiconductors — a phthalocyanine for the n-type elements and a thiophene for the p-type — on a flexible plastic substrate. The fanciest gadget they have made is a shift register, a ubiquitous digital circuit that shifts bits from a set of inputs to a row of outputs in an orderly fashion — microprocessors use them to do arithmetic and video cameras use them to scan pixels for incoming light. In the jargon



Dodabalapur: mixing the organic and inorganic.

of digital engineering, this circuit is approaching 'large-scale integration', meaning that a single chip contains thousands of components. For comparison, Intel's Pentium processors are now into the realm of ultralarge-scale integration, with almost ten million transistors etched onto a single chip. But the proponents of organic electronics note that many applications do not need this enormous processing power.

Not to be outdone, Philips has applied polymer electronics to display technology. The screen on a laptop personal computer looks crisp because it uses active-matrix technology, in which each pixel is controlled by its own transistor switch. In today's liquid-crystal screens, these are made from amorphous silicon. But the cost and limits to fabricating this silicon mean that it is not economically viable to build screens larger than about 38 cm across.

Philips has now created a crude active-matrix display using polymer transistors. Although this is a demonstration device measuring just 64 by 64 pixels, polymer electronics offers the possibility of building large yet affordable liquid-crystal displays.

De Leeuw's group at Philips also returned this month with an increase in the performance of their all-polymer integrated circuits⁴. The key was using materials with a higher

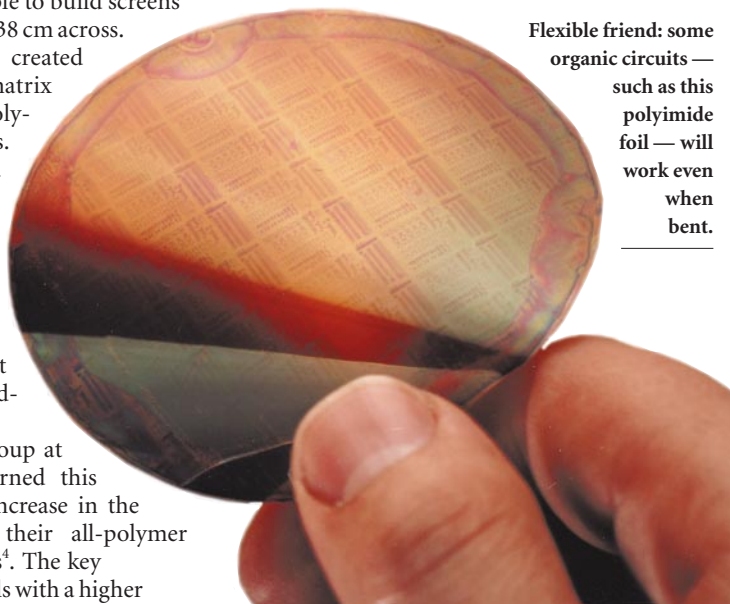
carrier mobility. This is an important attribute for any semiconducting material. Crudely put, mobility is a measure of how fast an electron or hole — the absence of an electron — can travel through a material. In practical terms it limits how quickly a transistor can switch on and off. Since even low-end electronics need to switch reasonably quickly, improving mobility is important. In their latest design, the Philips researchers use high-mobility organic semiconductors such as pentacene and hexylthiophene. So whereas the group's earlier circuits could only operate at switching rates up to about 200 hertz, the new versions operate at several kilohertz.

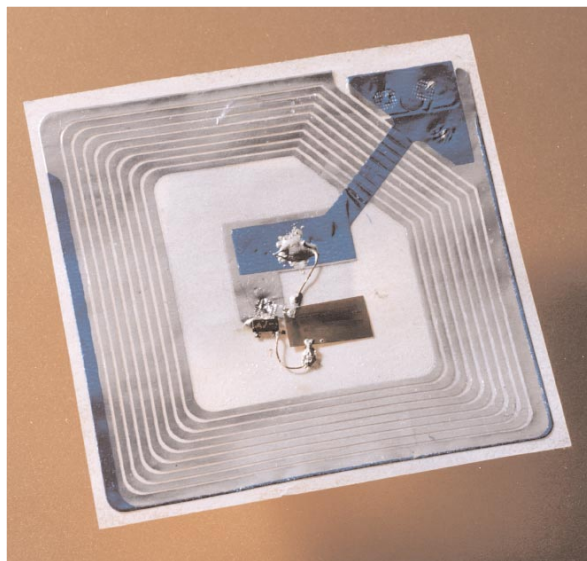
Upward mobility

The message here is that materials science is the key to the future development of organic electronics — find a better organic semiconductor and improved circuits will follow. Right now, the polymers that are easiest to build into electronic circuits have frustratingly low carrier mobility — about a million times lower than silicon. The problem is that there tends to be a trade-off between an organic molecule's mobility and the ease with which it can be processed. Because the molecular bonding in organic molecules is relatively weak, they tend to be mechanically flexible and soluble, which makes devices easy to fabricate by liquid deposition or the Philips group's spin-coating technique.

But the weak bonding also means that electrons hop from molecule to molecule relatively slowly. And attempts to boost mobility by strengthening bonding also increase the stiffness of the material and reduce its solubility. Organic molecules such as pentacene may have much better mobility, but devices constructed from them are more difficult to manufacture and more expensive. Other high-tech organic

Flexible friend: some organic circuits — such as this polyimide foil — will work even when bent.





In the shops now: integrated plastic circuits developed by de Leeuw (right) and his colleagues at Philips have found a use in anti-theft stickers for retail goods (above).



semiconductors degrade when exposed to the air.

Even so, those with experience of researching the properties of organic materials are confident that the problems can be solved. Cambridge's Friend believes that the key to improving mobility is to lay down organic molecules in ordered arrays, rather than spaghetti-like tangles. "You can arrange to get highly ordered structures if you have the right recognition between the substrate and polymer being laid down from solution," he says.

Studies of polythiophenes by Friend's group have already shown that highly ordered structures have improved mobility. The researchers deposited these polymers onto a substrate made of a mixture of silicon and silicon oxide so that the molecules organized themselves into ordered crystalline sheets either parallel or perpendicular to the surface. They discovered that the orientation affected carrier mobility, with this being two orders of magnitude higher in the parallel crystals than in the perpendicular⁵.

More recently, the Cambridge team has worked out how to lay down films of polyfluorene that assume a liquid-crystal state and become highly ordered⁶. This enhances the material's carrier mobility by a factor of about ten. And because much is already known about how to attain order within liquid crystals, this approach could provide a means to improve the carrier mobility of a range of organic semiconductors. "These are the sort of approaches that can be pushed much further," says Friend.

Another promising approach is to make hybrid materials — composites of organic and inorganic molecules that combine the benefits of each and, it is hoped, leave the

drawbacks behind. Christos Dimitrakopoulos and his group at IBM's Thomas J. Watson Research Center in Yorktown Heights, New York, have been studying molecular-scale composites that combine the properties of inorganic crystalline materials with those of organic molecules⁷. The idea is to retain the mobility of the inorganic components and make them easier to process by gluing them together, so to speak, with the organics.

The perovskite cookbook

The IBM team has been working with perovskites, in which inorganic molecules are linked together to form an extended layered framework. One class of perovskites has already attracted lots of attention as high-temperature superconductors, and it is well known that the materials' electronic properties can be tuned. Dimitrakopoulos and colleagues stirred together the inorganic compound tin iodide, which has a respectable carrier mobility, with the organic compound phenethyl ammonium. Then they spun-coated the mixture onto a substrate. As it dried, the composite material self-assembled to form multiple sheets of octahedral tin iodide molecules with organic layers sandwiched in between.

The resulting perovskite had a mobility of up to 0.6 square centimetres per volt second, six times higher than any other spin-coated material. Although not in the same class as silicon in terms of mobility, the layered perovskite beats silicon hands down in processability. Dimitrakopoulos expects that with some work, the material's carrier mobility might be pushed as high as 50 square centimetres per volt second, approaching that of pure tin iodide. Moreover, the perovskite could be fashioned into both p- and n-type

semiconductors — and so offers the hope of manufacturing low-cost complementary logic circuits.

Other researchers are improving the potential performance of organic electronic devices by concocting strongly insulating materials that can be laid down in very small spaces, which will prevent current leaking out of organic transistors and their electrodes. For example, Dominique Vuillaume and his group at the Northern Institute of Electronics and Microelectronics in Villeneuve d'Ascq, France, have been making extremely thin insulating layers for use in organic transistors. Lately, they have managed to self-assemble monolayers of alkyltrichlorosilanes just two nanometres thick that function as the gate insulator in a field-effect transistor⁸.

With materials scientists continuing to coax better performance out of organic electronic devices, the prospects for the field should continue to brighten. As yet, it is too early to predict the eventual scope of the plastic electronic revolution. But with Philips' anti-theft stickers already in use, and its active-matrix display at the prototype stage, the first of silicon's barricades have already been stormed. ■

David Voss is a science writer in Silver Spring, Maryland.

1. Garnier, F., Hajlaoui, R., Yassar, A. & Srivastava, P. *Science* **265**, 1684–1686 (1994).
2. Drury, C., Mutsaers, C., Hart, C., Matters, M. & de Leeuw, D. *Appl. Phys. Lett.* **73**, 108–110 (1998).
3. Crone, B. *et al. Nature* **403**, 521–523 (2000).
4. Gelinck, G., Geuns, T. & de Leeuw, D. *Appl. Phys. Lett.* **77**, 1487–1489 (2000).
5. Sirringhaus, H. *et al. Nature* **401**, 685–688 (1999).
6. Sirringhaus, H. *et al. Appl. Phys. Lett.* **77**, 406–408 (2000).
7. Kagan, C. R., Mitzi, D. B. & Dimitrakopoulos, C. D. *Science* **286**, 945–947 (1999).
8. Collet, J., Tharaud, O., Chapoton, A. & Vuillaume, D. *Appl. Phys. Lett.* **76**, 1941–1943 (2000).

Don't dismiss chlorine; it could help us to avoid the fate of the Romans

Sir — In his book *Pandora's Poison: On Chlorine, Health and a New Environmental Strategy*, recently reviewed by Terry Collins (*Nature* **406**, 17–18; 2000), Joe Thornton advocates a broad policy that would require industry to phase out chlorine-based technologies in favour of cleaner alternatives. He says we must do away with a regulatory system that looks at one chemical at a time, and replace it with a precautionary approach that addresses major classes of chemicals and industrial processes.

Considering all organochlorines, from essential non-toxic PVC to the very toxic 2,3,7,8-TCDD (dioxin), as one class to be discarded is as non-scientific as saying that all reptiles and birds are dangerous because some snakes are venomous.

There is no scientific basis for regulating all hydrocarbons — from non-toxic polyethylene (PE) to very toxic polycyclic aromatic hydrocarbons (PAHs) such as benzo(a)pyrene — as one class. Nor is there any scientific basis for regulating nitrogen-containing organics, from nylon to nitro-PAHs (the most powerful mutagens ever found, contained in diesel motor exhaust). Such a broad-based precautionary principle would mean that chlorine, nitrogen and every other organic material would be banned.

Chlorine is a very reactive element, making reactions possible that otherwise would be very difficult to perform without much more energy or more pollution. Chlorine is used in chains of processes: it is first used to make the building blocks for polyurethane and the remaining HCl is used to make PVC. That halves the energy needed to make chlorine for both. Without chlorine, the amount of (toxic) waste from polyurethane manufacturing would be much higher.

Thornton has been quoted in the online Environment News Service (<http://ens. lycos.com/ens/jun2000/2000L-06-08-06.html>) as warning that levels of dioxin in the environment can only increase, as long as organochlorines are produced. He claims that there is no safe way to dispose of them: "once they're in you, there's no way to get them out".

This is a clear untruth. In every Western country the levels of DDT/DDE, PCBs and dioxins in the environment and in animals and humans have fallen or are falling rapidly since these industrial products and by-products have been banned or restricted. PCB levels in North Sea fish are now half what they were ten years ago. The dioxin

content of Belgian mothers' milk has gone down by 30% in five years — mainly because of the stringent measures that were imposed on incinerators — during a period when chlorine and PVC production, use and incineration were higher than ever before.

"Like the Romans, who sipped from lead cups, ran drinking water through lead pipes, and bathed in lead basins, we have built our house of poison unaware of the consequences," Thornton says.

But unlike the Romans, who were likely to die by the age of 40 either from lead poisoning or as a result of infections, today's Westerners reach 80 years and older. In my view, this is not least thanks to drinking chlorinated water run through PVC pipes, swimming in chlorinated pools, eating food sealed in PVC-wrap foil, sleeping on mattresses made using chlorine and taking medicines of which 80% are made with the help of chlorine.

Ferdinand Engelbeen

Chlorophiles (an independent organization of workers in the chlorine/PVC industry), Oude Ertbrandstraat 12, B-2940 Stabroek, Belgium

Bright future in the stars for big telescopes?

Sir — We would like to correct a misleading impression left by your News article on the new Green Bank Telescope (GBT) (*Nature* **406**, 816; 2000), which achieved first light on 22 August 2000. Funding for the GBT was appropriated by Senator Robert Byrd (Democrat, West Virginia) not with an earmark, but with a \$75 million addition to the already established 1989 NSF budget.

While it is true, as your reporter wrote, that the project has taken considerably longer than initially planned, the telescope is not "tens of millions of dollars over budget", as it was built under a fixed-price contract. It is correct that claims by the contractor seeking additional payment are in arbitration, but your story appears to anticipate the arbitrator's decision while the hearing is still in progress.

More important, though, your article leaves the impression that the GBT stands alone as the "last of the giants", and that use of large single dishes is being superseded by other approaches.

On the contrary, there has been a resurgence in the use of single-dish telescopes, stimulated in part by astronomical discoveries, but also by the development of active reflecting surface technology and focal-plane array detectors. As well as the GBT, large single-dish telescopes are under construction in Sardinia and Mexico and one is planned in China. Major upgrades have been undertaken to instruments in

Puerto Rico, France, Germany and the United Kingdom. Single-dish telescopes are a critical component of world astronomical capability.

P. R. Jewell*, F. J. Lockman*, T. M. Bania†

**National Radio Astronomy Observatory, PO Box 2, Green Bank, West Virginia 24944, USA*

†Astronomy Department, Boston University, 725 Commonwealth Avenue, Boston, Massachusetts 02215, USA

Non-scientists thanked for vital help in Kansas

Sir — We appreciated your News story (*Nature* **406**, 552–553; 2000) about the role of scientists in restoring modern science standards to elementary and high-school classes in Kansas. We would like to point out, however, that our efforts, and the similar efforts of scientists in New Mexico, would have come to nought without the many non-scientists who were willing to devote hundreds of hours to this issue, and who were willing to drag us out of the laboratory.

The success of groups like Kansas Citizens for Science depends on the combined efforts of people with backgrounds in many fields, not just science, as well as support from national organizations such as the National Center for Science Education. When many are involved, the effort required from each scientist does not detract from research.

Despite the temporary debacle over science standards, Kansas is one of the highest-ranked US states in educational achievement. High expectations in science for high-school graduates can only make our job as educators and researchers easier. We urge readers to examine science standards in their area, then look for supportive non-scientists to help ward off attempts by creationists or other groups to bring religion into the science classroom.

Matthew Buechner

Department of Molecular Biosciences, 8035 Haworth Hall, University of Kansas, Lawrence, Kansas 66045-2106, USA

Other signatories to this letter:

Adrian Melott Department of Physics and Astronomy, University of Kansas

Robert Hagen Departments of Ecology and Evolutionary Biology, University of Kansas

Philip Baringer Department of Physics and Astronomy, University of Kansas

Patrick Ross Department of Biology, Southwestern College, Kansas

Keith M. Ashman Physics Department, Baker University, Kansas

Erratum

In a letter about journals (A. A. Waheed *Nature* **405**, 613; 2000), the third paragraph should start: "Just 12 countries each published more than 1% of the total: from 1966 [not 1996]–1998 their combined share averaged 88.9%."

Science-based security under threat

Los Alamos National Laboratory is an endangered national treasure.

David Pines

During the past year, scientists and the whole scientific environment at Los Alamos National Laboratory have suffered a series of staggering blows at the hands of the US Congress and the Department of Energy. Following the accusations of espionage against one of its weapons scientists, Wen Ho Lee, a US citizen of Taiwanese descent, the entire community of scientists at Los Alamos has experienced a series of restrictive measures that have seriously interfered with their ability to provide the kind of science-based security that has made the laboratory a national treasure during its 57 years of existence.

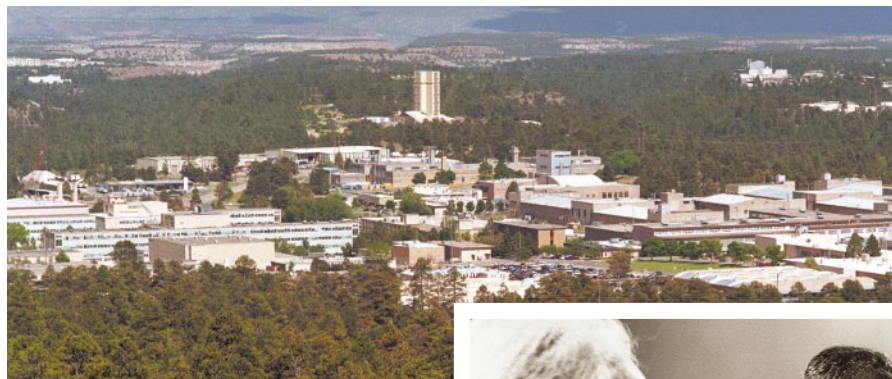
To cite one example, the 1999 congressionally mandated cuts in funding for vital research at Los Alamos led one of its main supporters in Congress, Senator Pete Domenici (Republican, New Mexico), to write to a colleague: "The House action is irresponsible ... the damage would be as serious and more assured than the suspected damage that may have been caused by Wen Ho Lee". These strong words accurately reflected the frustration and concern of scientists working at Los Alamos, concern that has only been heightened by the political and bureaucratic fallout from the recent apparent disappearance and reappearance of two disk drives containing information, not classified as secret, according to press accounts, on responding to nuclear weapons emergencies.

Value of Los Alamos

What Los Alamos has given the United States and its allies over its distinguished scientific and technological lifetime is a special kind of science-based security founded on four interlocking, essential ingredients: excellence, openness, leadership and oversight. All four were put in place in its earliest days; all four have been retained up to the present day; and, sadly, all four are now at risk.

The quality and dedication of the Los Alamos researchers working on the atomic bomb is legendary. But even the brilliant Robert Oppenheimer could not have recruited this remarkable team without being able to guarantee its members an open scientific environment that encouraged collaboration and was protected by the enlightened management of the University of California. Had either of these ingredients not been present, even under wartime conditions, Oppenheimer could not have assembled a team of the quality required for the laboratory's mission.

During the Cold War, the country turned



At risk: Los Alamos laboratory (above). Right, Oppenheimer in discussion with Einstein.

Meeting its commitments to national security depended on world leadership in science.

to Los Alamos for leadership in the development of nuclear weapons and weapons-supported research. Los Alamos supplied these at the same time as preparing to become a scientific laboratory devoted to pure as well as applied research. Thus, it was at Los Alamos that Frederick Reines and Clyde Cowan carried out the Nobel-prizewinning experiment in which neutrinos were detected directly for the first time.

Louis Rosen developed the Los Alamos Meson Physics Facility, where nuclear and condensed-matter physicists carried out state-of-the-art experiments on matter using the meson and neutron beams the facility produced. Peter Carruthers, with the encouragement of laboratory director Harold Agnew, hired a distinguished group of theoretical physicists, including Mitchell Feigenbaum, whose fundamental work on the dynamics of chaotic systems was to earn him a Wolf prize. It was at Los Alamos that the Human Genome Project was invented by the recently deceased George Bell and his postdoc Charles Delisi; while under the leadership of George Cowan, it was here that the Santa Fe Institute was born; and, based on its work on the actinide elements, Los Alamos developed a world-leading research group devoted to a fascinating new class of uranium-based materials. This led to the discovery by Zachary Fisk and Jim Smith of the heavy-electron superconductors



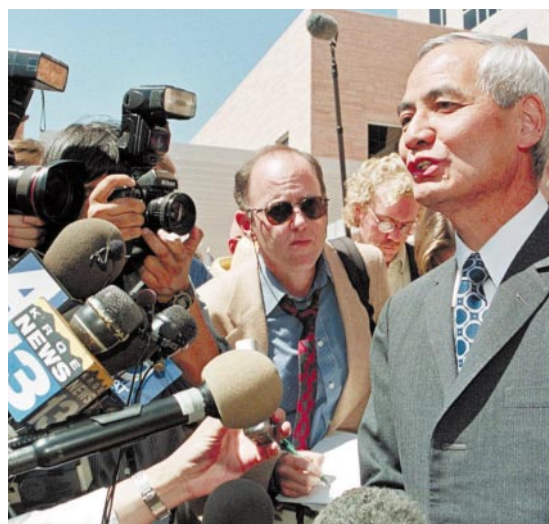
UPt₃ and (with Hans Ott) UBe₁₃.

The laboratory was well positioned to maintain its strength in the post-Cold War years, as its major mission began to change. It realized that meeting its commitments to national security depended on world leadership in science. It put in place the key ingredients for a successful science-based approach to the stewardship of the US nuclear stockpile and the reduction of nuclear and biological threats. Inside the security 'fence' where scientists carry out classified work, it had an open, scientific environment that was a fitting legacy of the Oppenheimer years, while outside the fence were groups working on fundamental problems whose solution could be important in enabling the laboratory to carry out its major missions. Hence, defence scientists had ready access to a pool of superb scientists, both within and outside the laboratory, who did not possess security clearance, but whose research results might be essential for the classified work going on inside the fence.

It is this effective integration of the defence science community into a world-class scientific environment that has made Los Alamos a unique source of strength for the security of the United States, and it is this unique strength that is now under attack by those in Washington DC who, with the laudable goal of improving the security of classified information, are proposing fundamental changes in that environment.

JOE RAEDLE/COLORIFIC

CORBIS



Freed speech: Wen Ho Lee faces the press after his release from nine months' imprisonment.

For much of the past 50 years, US weapons scientists had a compact with the country. In return for working in comparative isolation on applied problems of vital national interest, they were accorded the high degree of respect they deserved, and were freed from the micromanagement that goes on in industry and many government laboratories.

That compact has now been broken, I hope not irreversibly. Weapons scientists who have undergone a rigorous security screening and have worked for years at the laboratory are being told they can no longer be trusted with the secret information their work has generated, and that additional security measures, such as polygraph tests, must be imposed. Moreover, their work is being micromanaged by a bureaucracy located not only in Washington but also in field offices in Albuquerque, whose staff all too often have little or no understanding of the scientific issues involved, or the trade-off between total security and realistic working conditions.

Seismic shocks

At the same time, the fundamental scientific environment and the unique, large-scale scientific capabilities of Los Alamos, which have made it a scientific powerhouse for the nation, have suffered a series of seismic shocks, any of which would alone have produced significant cracks in scientific morale and productivity, but that taken together have brought about severe structural damage.

The actions began with a one-third reduction in the funds available to the Laboratory Directed Research and Development programme, an internally funded programme of carefully reviewed proposals from staff scientists, intended to lead to new directions in research. This reduction was mandated by a Congress apparently suspicious of providing funds for activities over which it had no direct control. It was accompanied by a 40% reduction in travel funds — the lifeblood of com-

munication for scientists working in a comparatively isolated environment — and followed by an effective cut-off of contacts at the laboratory with scientists from so-called sensitive countries — China, Russia, India, Pakistan and Israel.

Science is a truly international enterprise, and reducing the pool of available talent is obviously a poor idea. But despite claims that exceptions can be made, citizens of 'sensitive' countries can no longer visit the laboratory or attend workshops sponsored by the laboratory, much less be appointed as postdoctoral fellows or staff members.

How much damage is this doing? It is difficult to determine, but organizations are highly nonlinear systems. Changes in one part can have unexpected consequences in another.

Scientific renewal comes from the young, and from the outside, and for that reason Los Alamos had established an outstanding postdoctoral programme to attract excellent young researchers. Scientific renewal in the classified programmes at Los Alamos has typically come through an evolutionary process, as leaders of defence science research groups get to know, and then hire, promising postdoctoral researchers and other young temporary staff members. But perhaps as many as 40% of the best postdoctoral fellows in the past have come from these 'sensitive' countries. If the quality of this talent pool decreases appreciably, then over time the quality of defence science will also decrease.

With the end of the Cold War and at a time when the United States has so few obvious enemies, Los Alamos already faced major problems in competing with start-up companies and Wall Street for promising young scientists in many fields of great importance to national security. Not only has its competitive position now been further weakened by the externally imposed changes in its scientists' working conditions, but also members of the Asian community, suspecting that race was a factor in the Wen Ho Lee case, have encouraged an informal boycott of the weapons laboratories by US scientists of Asian origin. This has led to marked reductions in applications for postdoctoral positions by this highly talented community, which has in the past been a significant source of scientific strength for the laboratory.

There is one further threat, perhaps the most serious one, to the future effectiveness of the laboratory. That is the possible withdrawal of support by the Department of Energy for the oversight of Los Alamos by the University of California. By nearly all objective accounts, and certainly at the level of the working scientist, the university has always done a fine job of managing the laboratory. All the laboratory's employees are also employees of the University of California, with all the associated benefits. Moreover, a substantive collaborative effort between

members of the university faculty and the scientists at Los Alamos has been of great mutual benefit. If the university stops managing the laboratory, widespread resignations among the senior scientific staff would be inevitable. The laboratory's defence research would be especially hard-hit, as many of the scientific staff have been employed for more than 20 years, and would choose early retirement rather than start a new pension arrangement with a different employer.

Not too late?

None of these adverse developments is irreversible. There is still time for funding and policy changes to restore and improve the scientific environment at the laboratory. Los Alamos scientists are a remarkably hardy, loyal and optimistic group, so that despite all the above problems, there have been remarkably few resignations. Rather, scientists have continued their research, hoping and expecting that better times lie ahead.

To take an example of why I believe there are grounds for optimism, members of the screening committee for the Laboratory Directed Research and Development programme, on which I serve, agree unanimously that the quality of the proposals before the committee is significantly superior to proposals submitted last year, and may be the highest in the programme's history. I hope this is not a consequence of clarity of thinking induced by the approach of the hangman's noose, but rather a reflection of the extraordinary quality and dedication of the Los Alamos scientific community, which deserves so much better than it has been receiving at the hands of the media, the public, Department of Energy bureaucracy and Congress.

David Pines is co-director of the University of California's Institute for Complex Adaptive Matter and a staff member at the Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA. e-mail: pines@lanl.gov



Defence measures: armed guards sit outside the lab after the loss of two computer disk drives.

Taking the heat out of global warming

Two 'mainstream sceptics' take a critical look at the majority view.

The Satanic Gases: Clearing the Air about Global Warming

by Patrick J. Michaels & Robert C. Balling Jr
Cato Institute: 2000. 234 pp. \$19.95 (cloth); \$10.95, £7.99 (pbk)

Hans von Storch

Everybody knows about climate, and many people blame climate change and anthropogenic emissions for all kinds of weather-related catastrophes. In their book, Pat Michaels and Robert Balling accuse the current US president, and one of the candidates for the job, of misusing such fears in their policies. The authors argue that global warming is a scientific construct that owes its survival largely to the conservative reliance on a once-accepted paradigm rather than on rational arguments.

The authors' main motivation for publishing *The Satanic Gases* seems to have been the imminent US elections. They fear that the science of global warming could be used as a "basis for sweeping policy recommendations that could gravely harm United States prosperity". And they circle around US politics in a way that can be either amusing or annoying for a non-US reader. However, the book is more than this. Once the first parochial sections are dispensed with, one finds the authors' own view of the science of climate.

Michaels and Balling cover a broad range of subjects. They move from a short history of the Intergovernmental Panel on Climate Change (IPCC) to a description of the climate system, the observational climate record and the problems the IPCC has had in coming to grips with the fact that the recent warming has been smaller than the greenhouse theory would have suggested. They even make a quick excursion into the theory of scientific revolution. They give their own best guesses at future climate change and its impact in terms of sea level, mortality, vegetation and agricultural production. Both authors, defining themselves as 'mainstream sceptics', agree with the concept of global warming, but claim that it is at the lower end of the IPCC's spectrum of scenarios.

Their main argument is that, because contemporary climate models show a linear increase in temperature and other variables, the future increase should be a linear extension of these recent trends. Thus, they conclude: "The bottom line is a warming of around 1.5 °C to 1.7 °C in the winter half year and 1.2 °C to 1.3 °C in the summer over the course of the next century". Since, according to their analysis, winter cold inflicts far more harm than summer heat, they expect the

positive impacts of global warming to clearly outweigh the negative impacts. After reading a chapter explaining in some detail the various problems with general-circulation models, which concludes that they are suitable only for theoretical lecture-room studies, I wonder why they believe one particular feature of climate model simulations — the linearity of the temperature increase.

The authors' numerous references to the peer-reviewed literature tend to be biased towards publications giving estimates of a low climate sensitivity. Examples of this include scenarios of North Atlantic storms. And evidence in line with their adopted point of view is uncritically selected from the manifold of palaeoclimatic data. But as the book is a polemic, the evidence put forward seems reasonable, and the critiques raised make one wonder whether the summary of the IPCC Second Assessment Report was possibly also biased. Michaels and Balling find it particularly upsetting that, although the report was published in 1996, it used a crucial 'detection' paper that considered data only up to 1988, even though, after this time, warming ceased or was even followed by a period of cooling. They emphasize the as-yet unresolved inconsistency between

temperature trends at the Earth's surface and those in the mid-troposphere, where there has been no really significant warming in the past 20 years.

The book is not without errors. For instance, it claims that contemporary general-circulation models would not explain mid-latitude storms. And the attempt at a statistical explanation of the link between crop production and the concentration of air-borne carbon dioxide is based on the correlation of two trends, which would always return a high measure of similarity. But such a wide-ranging discussion could not be expected to be error-free.

Is the book worth reading? If you are only ever going to read one book on climate change, then no. Otherwise, it is useful to have a critical assessment of a majority view, and one which is — perhaps too early for its own good — set in stone in the IPCC reports. Climate research is no longer value free. When we questioned some 400 climate scientists from North America and Europe, most acknowledged this fact. Many climate scientists have become activists, promoting subjective policies for improving the world. Some want to do this by reducing greenhouse-gas and other emissions, others by



On an industrial footing

They began as postcards, a pictorial narrative of the America of the Vietnam era. *100 Boots* by performance artist Eleanor Antin (Running Press, £12.99) is a compilation of the 51

photographs of the boots on the beach, going to church, at the supermarket, at a job (detail above), and going to war. Their final resting place was New York's Museum of Modern Art.

book reviews

reducing the concern over a hypothetical future event. Both groups have non-scientific motives in their baggage. At least in Europe, the first group dominates, and it may indeed be very helpful to read what the other faction of scientists have to say.

The book's last section deals with the social process of science, how certain views, such as the probability that global warming will have disastrous effects, become dominating paradigms and how the organization of science may influence this process. These are relevant questions, and research agencies would do well to provide more funding for interdisciplinary research into such matters, particularly if they feed directly into policymaking.

The book is well written and easy to read. And the authors should be applauded for their bold predictions for the future, which will be proven right or wrong by the course of history: the Kyoto Protocol will have no effect; carbon dioxide emissions will continue to increase; by 2050, the Earth's average surface temperature will have increased by 0.65–0.75 °C in the winter half-year and by between 0.60 and 0.65 °C in summer; by 2050, crop yields will have risen sufficiently for the (carbon dioxide-related) rise alone to feed one-quarter of today's population; temperature-related mortality will decline.

Others make different forecasts. ■

Hans von Storch is at the Institute of Hydrophysics, GKSS Research Centre, Max-Planck-Strasse 1, PO Box, 21502 Geestacht, Germany.

To the core of consciousness

A Universe of Consciousness: How Matter Becomes Imagination/Consciousness: How Matter Becomes Imagination

by Gerald Edelman & Giulio Tononi
Basic Books/Allen Lane: 2000. 288 pp.
\$27.50/£20

Raymond J. Dolan

A Universe of Consciousness is the latest in a series of books in which the principal author, Gerald Edelman, develops a neurobiological theory of consciousness. It is no exaggeration to say that his previous books have aroused conflicting reactions. A frequent criticism has been that he couches fairly simple ideas in an idiosyncratic jargon that renders the end product impenetrable. Such criticism will be hard to level against the current volume, co-authored by Giulio Tononi, which represents the best and most accessible introduction to Edelman's ideas on consciousness.

A substantial portion of the book goes over familiar ground. In brief, Edelman has championed a view that brain function and organization are determined by Darwinian evolutionary principles that occur within the lifetime of the organism. Consequently, major sections are taken up with accounts of selectionism, the concept of value and the

critical role of neuronal interactions mediated by a process he has labelled re-entry. Re-entry is a much misunderstood idea, often confused with feedback. It refers to the dynamic interchange between reciprocally connected brain regions that is fundamental to synchronization and the coordination of their mutual functions.

All of these ideas are treated clearly and concisely. There is also wide acknowledgement of competing ideas on the structure of consciousness and, although it is clear that the main issue is higher-order consciousness, there is an acceptance that a complete account will include what has been termed primary consciousness.

A Universe of Consciousness still provides its moments of frustration. A prime example is the treatment of memory, in which the authors propose the controversial idea that all forms of memory are non-representational — here the authors provide the useful geological metaphor that memory is more like the melting and freezing of a glacier (non-representational) than an inscription on a rock (representational). In their account, memory results from selective matching between distributed neural activity and signals from the world, the body and the brain itself. But this potentially radical proposal is not developed sufficiently.

The book's novel contribution is the development of what is termed the "dynamic core hypothesis". Put simply, a dynamic core is a subset of brain regions identified by the pattern of neuronal interactions within the core and its associated regions. As an analogy, the core can be considered akin to a shoal of fish. The envelope, or boundary, is clearly discernible but continuously changes as the fish (neuronal activity) swim to different locations (different neuroanatomical domains). The fish swim in concert, but each has its own unique trajectory through the water (that is, neuronal activity is coherent but not identical).

Critically, the core is characterized by being functionally disconnected from the rest of the brain. This disconnection is represented by relatively independent neuronal dynamics inside and outside the core. Thus, the authors liken the dynamic core to a functional cluster. The second key idea is that patterns of activity within the core exhibit a high degree of complexity. The complexity of the core is important because it reflects a high degree of differentiation, implicit in observing a particular brain state, given the diversity of patterns that could be expressed. The boundaries that define the core are predicated on constantly changing patterns of activity and are themselves dynamic, reflecting the functional state of the brain.

The notion of a dynamic core is both conceptually appealing and usefully grounded in information theory. There is, however, a problem with it. This arises from the fact that



Man made machine?

Is it possible that we will eventually become robots ourselves, uploading our consciousness into computers? *Robo Sapiens: Evolution of a*

New Species by Peter Menzel and Faith D'Aluisio (MIT Press, \$29.95, £19.95) shows how far we have come along that path.

it blends the theory of dynamical systems with information theory relating to stationary stochastic systems. The problem can be seen by considering a dynamic core over a short period of time. The complexity within this time frame is represented by the shape of the probability distribution of the activities of the elements comprising the core. The question is: what does this probability distribution refer to? Does it refer to realizations of brain activity patterns over time or to different realizations at the particular time that that core exists?

The authors have assumed the former when measuring complexity using neuro-imaging time series. However, it cannot be applied to the dynamic core hypothesis because any particular core is defined only over a short period of time. If the core differentiates between a vast repertoire of potential states by expressing just one state during its transient existence, then it cannot be complex, because the repertoire of states within its lifetime must, by definition, be severely restricted. If it were not restricted, no differentiation would occur and there would be no information about which brain state had been selected.

The second possibility, that complexity refers to realizations at the same point in time, is more tenable. And it could be linked to work on dynamic correlations that employs multiple trials to look at the correlations among neurons as a function of peristimulus time. However, this formulation precludes the existence of a unique dynamic core that exists only once in the lifetime of the organism (that is, conscious states must be reiterated). In short, by basing the definition of a dynamic core on the probability distribution of its underlying activity the authors imply that there is a repertoire of potential realizations that engender this probability distribution. Because the dynamic core is transient, with continuously changing boundaries, these realizations cannot be over time. Therefore, unless the same dynamic core is recapitulated exactly, on many different occasions, the probability distribution, on which its definition rests, does not exist.

Despite this criticism, the dynamic core hypothesis is a highly compelling idea. One can easily see how such ideas would flourish in the context of dynamical systems theory, where the creation of information and measurement of entropy in nonlinear dynamical systems are natural consequences of sensitivity to initial conditions.

The authors' proposal that consciousness is based on distributed neural activity necessarily rules out any attempt to assign a different neural group to every phenomenological state. When translated into neuronal terms, this enterprise rapidly runs up against a problem of scale, given that the approach requires that every property is represented by a distinct neural group. The question posed by the authors is: why should we suppose that firing

in any particular group of neurons is associated with a specific subjective quality? In their own words, "the enormous *petitio principii* so unveiled ... would trigger a sudden fit of inextinguishable laughter in the Olympians".

Explaining consciousness has become the Holy Grail of modern neuroscience. Any reckoning on who has found the true path is surely premature. Nevertheless, the account of consciousness provided by Edelman and Tononi is certainly highly plausible and can be recommended as one of the most ambitious accounts around. ■

Raymond J. Dolan is in the Wellcome Department of Cognitive Neurology, Institute of Neurology, London WC1N 3BG, UK.

..... A dip in the soup, with a pinch of salt

The Spark of Life: Darwin and the Primeval Soup

by Christopher Wills & Jeffrey Bada
Perseus: 2000. 288 pp. \$27, £16.50

William J. Hagan

Conventional images of the early Earth portray a hostile planet with violent displays of lightning and volcanic eruptions. To these pyrotechnics, one might add the occasional asteroid impact and the smoky plumes of hydrothermal vents. Beyond providing melodrama for artistic depictions (including those reproduced in this book by Bill Hartmann of the University of Arizona), such phenomena suggest possible energy sources that may have played a role in the origin of life. In *The Spark of Life* Christopher Wills and Jeffrey Bada review our understanding of the origins of life for the informed general reader.

There has been a lot of interest in evaluating the relative importance of different routes towards the organic precursors of the first life. One strategy is to assess the fluxes of energy on the modern Earth, then to estimate the evolution of these energy sources through time, and finally, to determine the relative rates of formation and decomposition of potentially prebiotic compounds. Although the first task should be the most straightforward, the compilation of energy inputs for the modern Earth has been complicated by shortage of data and by propagation of values from the older literature. Chris Chyba and Carl Sagan pointed out in 1991, for example, that the rate of terrestrial lightning dissipation is about 20-fold lower than previous estimates.

Controversy (often acrimonious) has also centred on the role of hydrothermal vents in prebiotic synthesis, since their discovery two decades ago. Although the large temperature gradients and hydrogen-rich conditions make them attractive to those (called "ven-



Murky birth? 'Ventists' believe that the precursors to life may have originated in hydrothermal vents.

tists" in this book) who believe that vents played a major role, critics, including one of the authors, have stressed the sterilizing effect of marine cycling through these hot regions. Agreement has been hampered both by the inaccessibility of these deep environments, and by the difficulties of laboratory simulation. The latter challenge is being addressed by several research groups, whose data it is hoped will confirm the results of thermodynamic calculations. Even if vents did facilitate the formation of small biomolecules, the transient lifetime of particular smokers makes it unlikely that a sustained sequence of reactions leading to protobionts could occur there; these organic sources may have simply augmented the overall concentration of the primordial soup.

The role of sunlight in prebiotic synthesis has become less fashionable, but any ranking of energy sources always puts ultraviolet and visible radiation at the top. Calculating the distribution of sunlight on the early Earth is complicated by the fact that, during the Sun's T-Tauri stage 4.6 billion years ago, its output was about 30% dimmer and ultraviolet light was more intense. Nevertheless, the total flux at wavelengths less than 230 nanometres 4.2 billion years ago is comparable to that reaching the edge of the Earth's atmosphere today, and this energy could have played a role in the formation of small precursors such as formaldehyde. Although longer wavelengths would probably have penetrated to the surface (especially in the absence of an ozone shield), solar-driven processes in this spectral regime have been less extensively investigated in the context of the origins of life.

The Spark of Life takes a sceptical view of

current models for the molecular precursors of life. As an example, the transport of organic materials to the early Earth by interplanetary dust particles, for which frictional heating would be small, yields the 'bad news' that only one microgram per litre of amino acids would be dissolved in the primordial ocean. Yet such calculations clearly depend not only on the delivery rate, but also on the (highly variable) half-lives of the organic compounds in these micrometeorites. At present, we lack enough analytical data to inform such estimates, which could yield much higher concentrations of robust molecules. Moreover, a dilute soup would be advantageous in its protection from photochemical decomposition, and most workers agree that some adsorption and/or partitioning process would have been essential for subsequent catalysis.

The authors provide an excellent and compellingly written overview of the spontaneous-generation controversy and other historical themes as they relate to the emergence of life. The coverage is not comprehensive, but is lively and includes a sampling of topical approaches, from meteorite analysis to the study of subsurface microorganisms.

Neither author is an organic chemist: Wills is a geneticist at the University of California at San Diego, and Bada is a geochemist and exobiologist at the Scripps Institution of Oceanography at La Jolla, California. This background is evident in occasional lapses, such as the claim that "RNA, which probably appeared before DNA, uses the more stable molecule uracil instead of cytosine [thymine?]." (For it is thymine in DNA, not cytosine, that forms less stable adducts than uracil — as occurs during photohydration — which could explain this base change when DNA arose in evolution.) Errors of attribution arise in the discussion of Leslie Orgel's 'life on the rocks' scenario for the synthesis of RNA oligonucleotides — although Orgel did develop the theory behind this process, he relied on the explicit collaboration of James P. Ferris, whose laboratory has pioneered the study of clay-catalysed oligomerizations of activated nucleotides. Although a popular account may be allowed some licence in the assignment of credit, the authors are notably more generous to co-workers when they are indigenous to southern California.

Nevertheless, this book provides a highly readable survey of the historical prelude to the study of the origins of life, as well as selected areas of current research, including the search for extraterrestrial life. ■

William J. Hagan is at the College of Saint Rose, 432 Western Avenue, Albany, New York 12203-1490, USA and the New York Center for Studies on the Origins of Life, Cogswell Laboratory, Rensselaer Polytechnic Institute, Troy, New York 12180-3590, USA.

Pitfalls, blind alleys and much more

Through the Rearview Mirror: Historical Reflections on Psychology

by John Macnamara

MIT Press: 1999. 266 pp. \$39.95, £26.50

Sam Glucksberg

John Macnamara taught psychology and the history of psychology at McGill University until his death in 1995. For those who knew him, his reflections on the history of psychology will evoke memories of his intense personal approach to his work. For those who did not know John, his affectionate survey of the interplay of psychology with philosophy and Christianity will introduce a lively intellect striving to understand how conceptions of human nature have developed over the past 2,000 years of Western thought.

Macnamara begins with what some might see as an invidious comparison between psychology and physics. He argues that, whereas Aristotle's *Physics* may be comfortably ignored by contemporary physicists, his psychology (in *De anima*; 'On the Soul') would amply repay careful reading. Why? Because Aristotle, Plato and other great figures in Western thought not only place modern ideas in historical perspective, but would also, if taken seriously, alert contemporary thinkers to "foreseeable pitfalls and blind alleys". A case in point is Thomas Aquinas' *De ente et essentia* ('On Being and Essence'), which, according to Macnamara, "refuted the modern theory" of concepts as abstractions from perception (emphasis added).

Macnamara's view of psychology and its history is, in a very real sense, parochial. It is explicitly limited to Western psychology, and excludes conceptions of human nature in any other of the world's cultural and religious traditions. Moreover, it is explicitly limited to a narrow, though central, set of human capacities: beliefs and belief-informed desires. And finally, it is concerned with the interplay of philosophy and psychology with Christianity.

Surprisingly for a scholar of such broad interests, Macnamara barely touches on the interplay of psychological ideas with the social, political and economic issues of the times. He begins by concentrating on conceptions of learning and truth, and how people might acquire and represent these concepts. He shows that the contemporary tension between nativists, who argue that ideas may be innate rather than acquired through experience, and empiricists, who claim experiential bases for human knowledge, has surfaced repeatedly in the history of ideas, from Aristotle and Plato to the radical behaviorism of John Watson.

Curiously, he does not bring his historical



Lessons from the master: could Aristotle's *De anima* be of help to today's psychologists?

survey to the present, where such thinkers as Noam Chomsky continue to argue for innate ideas and idealized (symbolic) representation against modern associationists who use connectionist networks to extract generalizations from experience. But the lessons of history, traced through 27 chapters, make it clear that some form of this debate has captured the attention of the major figures in philosophy, biology and psychology for more than 2,000 years.

And yes, even in theology. Beginning with Aristotle, and ending with the gestalt psychologists of the 1940s, who emphasized the holistic nature of perception, Macnamara takes us through the major movements in philosophy and psychology, along with a side trip to the Book of Genesis, an essay on the impact of Christianity on psychology and, in three chapters, a summary of how St Augustine and Thomas Aquinas influenced Western psychology and philosophy.

Surely, a 266-page book with 27 chapters, each a self-contained vignette, must represent the epitome of superficiality. But this would miss the point of Macnamara's argument: that even a nodding acquaintance with the ideas and debates of the past will put contemporary debates on the nature of human knowledge into perspective. He firmly believed that people who are ignorant of history are condemned to repeat it. Unfortunately, even those who are immersed in history may still be condemned to repeat it. I seriously doubt that even the most careful reading of the history of psychology would save contemporary workers from pitfalls and blind alleys. Even more to the point, it would neither prevent nor resolve the conflict between modern nativists and empiricists. ■

Sam Glucksberg is in the Department of Psychology, Princeton University, Princeton, New Jersey 08544, USA.

An electromagnetic personality

An Elizabethan experimentalist and spirited foe of pseudoscience.

James D. Livingston

Amidst all the millennial media hoopla, one significant anniversary has passed unnoticed. This year is the quadricentennial of the publication of William Gilbert's masterpiece, *De Magnete*. Galileo praised it lavishly, Kepler cited it as an influence on his work on planetary motion, and to the science historian George Sarton it was "one of the greatest books of Science".

William Gilbert was born in Colchester in 1544 and educated at Cambridge. He practised medicine in London, and held various positions in the Royal College of Physicians, including president. In 1600, Gilbert was appointed one of Queen Elizabeth's personal physicians. On her death in 1603, he became physician to her successor, James I, but Gilbert outlived Elizabeth by only a few months. Although his brother later published some of his other writings, Gilbert's fame as the founder of magnetic and electrical science rests entirely on *De Magnete*. Dryden memorialized him with this couplet, of which at least one of the conditions still ensures his immortality: "Gilbert shall live till lodestones cease to draw/ Or British fleets the boundless ocean awe."

Gilbert argued that "stronger reasons are obtained from sure experiments and demonstrated arguments than from probable conjectures and the opinions of philosophical speculators". He then proceeded to examine, by careful experiment, a wide variety of common beliefs about magnets and magnetic forces. Previous authors had claimed that magnets lose their attractive power when rubbed with garlic, or in the presence of diamonds. He showed otherwise, and ridiculed many other legends, including claims that a lodestone placed under the head of a sleeping woman would drive her out of bed if she were an adulteress — although in this case, his experiments went undescribed.

Earlier authors had lumped together the mysterious powers of lodestones and amber. But Gilbert's experiments showed the differences between magnetic forces, which were limited to lodestones and iron, and electrostatic forces, which could be produced by, and act on, many different materials. These tests were the first to demonstrate electrostatic forces using an electroscope, and in describing his results he coined the term 'electric', based on the Greek word for amber. There was little progress in the science of electricity and magnetism beyond *De Magnete* for nearly two centuries, until Coulomb's more quantitative measurements.

Among Gilbert's most important experiments were those studying the orientation of iron needles close to a spherical lodestone. Recognizing the link to mariners' observations of the dip of compass needles, he concluded "*magnus magnes ipse est globus terrestris*" (the Earth's globe itself is a great magnet) — a remarkably wide-ranging conclusion reached from simple table-top experiments.

Ocean navigation was of great importance in Elizabethan England, and Gilbert described the variation of compass direction and dip around the globe in great detail, with the ultimately unsuccessful goal of using compass readings to determine both latitude and longitude. As his experiments had shown that opposite magnetic poles attract, the south pole of the compass needle obviously pointed north, and anyone who defined magnetic poles differently was in error, "so ill-cultivated is the whole philosophy of the magnet still, even as regards its elementary principles". Gilbert's logic was impeccable, but failed to convince compass users to reverse their labelling. Thus by standard scientific

terminology, the Earth's south magnetic pole is in the Arctic, and its north magnetic pole in the Antarctic, but don't blame Gilbert.

Much of *De Magnete* is a ground-breaking demonstration of experimental physics, but in the final sections Gilbert extended his thinking into theories of astronomy and cosmology. These metaphysical excursions drew strong criticism from Francis Bacon, whose name we associate with the scientific method. But in Sarton's judgement, Bacon himself "would have been unable to apply his own method as brilliantly as Gilbert did".

An inexpensive paperback edition of *De Magnete* is available in English translation, and parts of it are still very lively reading, particularly where Gilbert fumes about writers who promulgate untested legends as fact. About reports of magnet-based perpetual-motion engines: "May the gods damn all such sham, pilfered, distorted works, which do but muddle the minds of students!" About broad medical claims for magnets: "Thus do pretenders to science vainly and preposterously seek for remedies, ignorant of the true causes of things." Four centuries on, perpetual-motion machines based on magnets are still promoted, and 'therapeutic' magnets are sold by the millions to a gullible public. Gilbert would still be angry, but I don't think he would be surprised. ■

James D. Livingston is in the Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 12309, USA.

He ridiculed claims that a lodestone would drive a sleeping adulteress out of bed.



A. Ackland Hunt's (probably imaginary) painting of Gilbert showing his experiments to Elizabeth I.

Through the shaving mirror

...or, A tight night at the Surrealist Sporting Club.

Michael Moorcock, after Maurice Richardson

"Has anyone noticed," Monsignor Cornelius spoke urgently, hoping to divert Sir Perkin Float, on his third bottle of claret, from developing that familiar litany about discovering Chaos Math years before Mandelbrot, thus being cheated of his place in history and his video royalties, "how cats can turn off time? With a suitable lens, of course."

Engelbrecht, the dwarf metaphysical boxer, grew alert. Professional curiosity. A founder of the Surrealist Sporting Club, he refused to fight anything lighter than a cathedral clock. Against advice he'd challenged Big Ben to a 'no quarter' fin-de-millennium celebratory bout. Serious Soho backers. Chinese calendar promoters. That the parliamentary clock had accepted was surprising, that it lost was suspicious. Strong rumbles in the sporting fancy. Someone had slipped the monster timepiece a heavy envelope to lie down. London was now on Chinese time, with all serious punters refusing mah-jong bets involving politicians.

"It reminds me," said the time-battered pug, "of that night in New York I almost lost to the Union Square Clock Tower. My career would have been over if it hadn't been for some fancy photon-work."

Tactfully the Corinthian Jesuit drew us into Engelbrecht's confidence. "Your mother discovered that time isn't a dimension of space, but a field whose properties are affected by the nature of space existing within it?"

"Space a quality of time?" Sir Perkin snorted into his wine. Glinting rubies fell to the linen. But Cornelius's clever sourcing meant outright disagreement would be dangerous.

Clearing the cloth, Engelbrecht used a carpenter's pencil and the condensed mathematical logic developed at his famous Mar-rakesh asram to illuminate us. "Time alters when it interacts with space. In common with all observable nature, the Universe, or multiverse, grows organically and is best imagined as a vast tree, or perhaps even a forest with common roots."

"And the soil for this tree?" Float's reckless scepticism terrified us. Expensive watches would only be the first victims of our dwarf's distemper.

The tiny slugger observed philosophically that this was the level of logic he must commonly suffer. "An analogy," he growled. "There's a theory that the multiverse is creat-

ed by the common will, but as to its origins..." He cracked his knuckles. "I think therefore I thump."

Float's timid attention returned to the claret. The merest whisper of Big Bangs had him reaching for his jug.

Engelbrecht scowled reminiscently. "We're familiar with the disappearing neutron, we've recently learned how light can travel faster than light. Conventional method produces Heath Robinson physics turned into formulae by crazed Euclidians. At some point, as Columbus told the Pope, we have to let go of the premise that the world is flat."

"Can we see these alternative worlds?" 'Prof.' Aspinall had been kicking the gong around and wasn't ready for further shocks.

Cornelius embellished smoothly. "I understand it's a question of scale and mass. Put simply, millions of subtly different versions of our reality are separated by size. Each version, though scarcely different in terms of the multiversal compass, is as invisible to us as if we were only seeing a single magnified pixel out of a complex computer image. We never see the whole. It is either too small or too large. We coexist in the same space through scale. Each alternative world has

greater or lesser density and is invisible to the others."

"Proliferating to infinity. Whales, fairies," Engelbrecht paused, "dwarves."

"Space curves," mumbled 'Prof.' into his spoon. "Don't it?"

"Organically and often, as a branch curves." Cornelius smiled. "Like ourselves, space consists of spheres, but isn't itself spherical. Nature would be contradicted if it were. Certain entities somehow adjust their mass and move 'intrabranally'. For instance, few creatures are as expert at varying their size as cats. Thus their mysterious 'disappearances'. Happily the phenomenon hasn't occurred with dogs."

"Bigfoot, however..." 'Prof.' began.

"Cats," said Cornelius hastily, "see space invisible to us, coming and going through the multiverse pretty much at will."

"Bunkum," hiccuped Float. "We were supposed to hear about that New York match our dwarf won by a whisker. In funny circumstances."

Engelbrecht swelled. "A classy clock fighting for a consortium of Istanbul high rollers backing the Julian calendar. Twice my form and landing some tricky byzantine jabs. By Round 60 I'd borrowed all available time. I'm on my back looking like someone just unwrapped the Mummy."

"I've already squared the ref, of course. While shaving I've also dreamed up some insurance. Fortunately I have a prism hidden in one glove, a photon in the other. Resting on the count with my hands invisible I pull Harness's old baffled quantum trick. On nine. Works like a charm. Time hesitates. My seconds Coleman and Benford produce the mirror, rescaling mass and size to shrink the heavy bastard enough so I can stagger up and deliver the dynamite. Gravity completes the job. Down he goes. Dead weight. Space-time readjusts. It's all over. The Yankee Boomer's stretched full size on the canvas, his gob-smacked hands chasing themselves round his face. Cheering punters. GMT keeps the title!"

"Convincing," admitted Aspinall. "Except no way would Coleman and Benford help snatch a fix."

Engelbrecht winked. "Prof., you can't name a physicist in the multiverse who isn't in my pocket. Now, padre, your cats?"

"Oh, another time I think," said Cornelius, contentedly filling his pipe.

Float was at last profoundly asleep. ■

Maurice Richardson's *Exploits of Engelbrecht* is republished by Savoy. Michael Moorcock's *Multiverse* is published by DC Comics.



Evolving evolvability

Linda Partridge and Nicholas H. Barton

In yeast, a modified protein known as a prion generates variation in growth rate across diverse environments. Is this an example of an agent that has evolved in order to promote its possessor's adaptability?

The mechanism of natural selection is well understood for genes that affect the survival and reproductive success of their bearers. Selection can also act on genes that likewise affect their bearer's relatives, or (less effectively) on the group of which the bearer is a member. But can selection favour genes or other factors that help their bearer adapt to changing circumstances? That is, can evolvability itself evolve? Such genes would be subject to a different, much weaker, kind of selection.

There is nonetheless the possibility that modifying factors may increase the rate of mutation or of genetic recombination (the shuffling of genes that occurs during sexual reproduction). Such factors can increase in frequency if they can hitch-hike by association with the new, advantageous mutations¹ or gene combinations² that they create. There are two problems, however. First, most new mutations and gene combinations lower reproductive success rather than increase it. Second, if reproduction is sexual, the modifier will tend to be parted by recombination from the novelty it has caused.

So much for the problems. What evidence do we have that such modifying factors exist? One such mechanism for generating variability has been brought to light by True and Lindquist, as they describe on page 477 of this issue³. They have been working with a yeast prion that they suggest is maintained by natural selection because it aids evolution.

The prion, called $[PSI^+]$, is a self-propagating and inherited modification of a protein, Sup35, that is involved in terminating protein translation — the process by which proteins are manufactured by translation of the messenger RNA codons into an amino-acid sequence (Fig. 1a). The prion occurs naturally, and yeast cells switch spontaneously back and forth between the prion and wild-type (normal) conformation of the protein. The prion form of Sup35 makes inactive 'termination complexes' that can capture newly made, wild-type Sup35 protein and convert it into the prion form. The prion causes 'read-through' of experimentally introduced stop codons, and presumably also of some naturally occurring stop codons. Because these codons signal a halt to translation, read-through of them results in proteins with an extra segment (Fig. 1b).

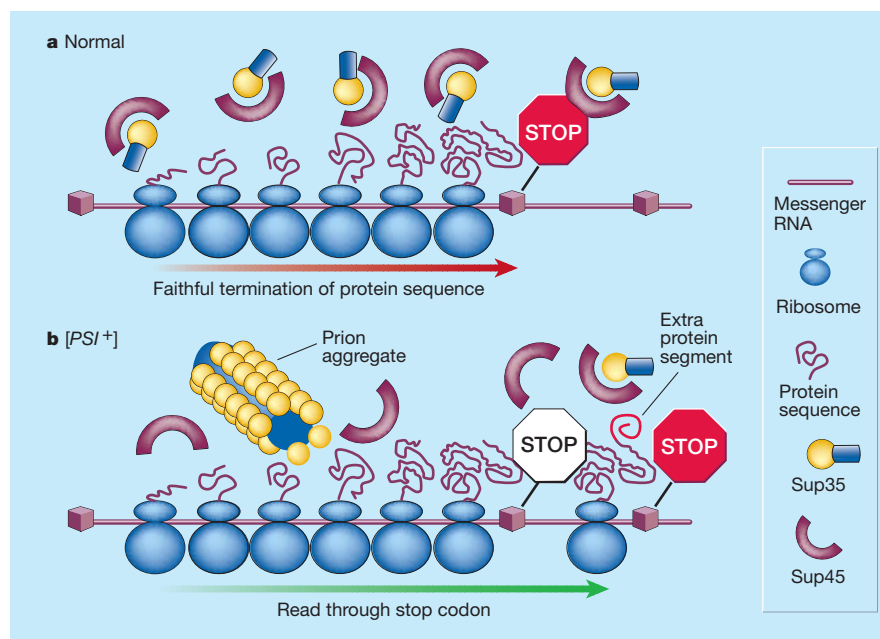


Figure 1 Protein production and the effects of the Sup35 protein and its modified form, the $[PSI^+]$ prion, in yeast. a, In cells with normal Sup35 protein, translation of the messenger RNA sequence into a protein sequence by the ribosome is normally terminated by a stop codon. Termination requires an interaction between Sup35 and other proteins such as Sup45. b, In $[PSI^+]$ cells, the Sup35 protein takes up the prion conformation and forms aggregates that fail to interact properly with the termination complex. Stop codons are therefore sometimes missed, producing proteins with an extra segment. True and Lindquist³ argue that the prion is maintained by natural selection because the variant proteins it induces sometimes increase yeast-cell ability to survive environmental change. (Graphic modified from material provided by Julie Feder.)

In principle, this provides the potential for evolutionary change.

One end of Sup35 is essential for its role in terminating translation and for survival of the cell. The other end and middle (N and M) regions are required for the protein to assume the prion conformation. If these regions are deleted in cells with the prion, the prion conformation is lost. But there are no other gross effects, such as death of the cell, raising the issue of how these regions are maintained by selection.

True and Lindquist³ tested the idea that the prion introduces novel characteristics into yeast cells. They measured the growth rate of yeast strains in different environments, such as on different carbon and nitrogen sources. They used a set of pairs of strains that differed from one another only in the presence of wild-type or prion protein conformation, while each pair of strains differed in genetic background and in the extent of read-through of stop codons. In nearly half

of 150 tests, the prion altered the growth rate, increasing it in more than a quarter of them. The effects on growth were specific to particular strains and environments. These findings lead True and Lindquist to suggest that the N and M regions of Sup35 might be maintained by selection because, under conditions that are inhospitable to wild-type yeasts, cells that spontaneously switch to the prion state may by chance fare better and increase in frequency.

This is an intriguing proposal, but is open to debate. The observation that deletion of the N and M regions of Sup35 has no obvious effect, apart from a lost ability to take up the prion conformation, does not necessarily mean that they have no other function. After all, genes can often be deleted with no serious effects under laboratory conditions⁴, and yet are maintained by selection in nature. True and Lindquist carried out a more direct test of other possible functions of the N and M regions by measuring growth rates of strains

with deletions for these regions. These deletions did influence growth, so — as the authors acknowledge — the N and M regions may well have effects beyond those mediated by their ability to take up the prion conformation.

If selection for 'evolvability' is to maintain the N and M regions of Sup35, then the prion version of the protein must increase in frequency because it occasionally creates favourable variants. Even if the chance of stumbling on such a variant is low, the consequent increase may be so large that genomes carrying the N and M regions increase on average relative to those that do not. Increased mutation or recombination rates can be favoured in the same way: a gene that modifies the genetic system gains an advantage by associating with the favourable variations that it generates. A peculiar feature of True and Lindquist's example is that variants are produced by a heritable change in protein conformation, rather than by a genetic change. So if the environment changes back, the [*PSI*⁺] strain can spontaneously revert to the original type. Similarly, Taddei *et al.*⁵ showed that genes that raise mutation rates can hitch-hike with the favourable variants that they produce, and yet may not become fixed in the genome: back-mutation causes reversion to the original low-mutation genotype, and hence avoids the harmful consequences of permanently increased mutability. Such reversion makes selection for 'evolvability' more plausible.

These results³ show that previously hidden variation may be revealed by perturbations to the mechanism of gene expression. Other examples are well known, and have been taken as evidence that selection has shaped the organism so as to direct development towards the optimal state^{6–8}. As True and Lindquist point out, and as many experiments have demonstrated^{7,9}, variants revealed by non-genetic perturbations may be picked up by selection and thus permanently incorporated into the genome. However, we feel that it is simpler to see the increased variability in these examples as a side effect of disrupted gene expression, rather than as an adaptation to facilitate evolution.

True and Lindquist's study is motivated by the idea that evolution requires "the concerted effects of several independent genetic changes", each of which is "deleterious individually"; indeed, this apparent difficulty is a frequent criticism of darwinian evolution. But it is not, as the authors state, "A major enigma in evolutionary biology": there is no evidence that the individual steps towards a complex adaptation are deleterious¹⁰. The final product may be complex and finely tuned, such that most changes are deleterious, but that does not mean that there is no path by which natural selection could construct the adaptation. The power of natural selection is that it assembles a series of

changes, each individually tested; mechanisms that produce large variations, involving several random changes, are unlikely to be helpful. So we disagree with the authors over whether new mechanisms for generating radically different variants would help explain adaptive evolution.

Nonetheless, True and Lindquist's hypothesis is feasible in principle, and it is testable. To demonstrate that the N and M regions of the Sup35 protein are maintained by occasional selection for the evolutionary novelties produced by the prion, one could look in nature for a correlation between the presence of the prion and stressful environments. More practicably, one could allow laboratory strains with and without the N and M regions of Sup35 to compete in both constant and varied environments.

It is likely that specific gene sequences have evolved to generate specific kinds of variation¹¹ (for example, in the vertebrate immune system). Moreover, the genetic recombination that occurs during sexual reproduction is likely to have evolved as

an adaptation to generate variability. Yet it is sobering that, despite several decades of intense research, we still have little direct evidence that genetic systems have in fact evolved to facilitate evolution². ■

Linda Partridge is in the Department of Biology, University College London, Wolfson House, 4 Stephenson Way, London NW1 2HE, UK.
e-mail: ucbhlop@ucl.ac.uk

Nicholas H. Barton is in the Institute for Cell, Animal and Population Biology, University of Edinburgh, West Mains Road, Edinburgh EH9 3JT, UK.

e-mail: n.barton@ed.ac.uk

1. Johnson, T. *Genetics* **151**, 1621–1631 (1999).
2. Barton, N. H. & Charlesworth, B. *Science* **281**, 1986–1990 (1998).
3. True, H. L. & Lindquist, S. L. *Nature* **407**, 477–483 (2000).
4. Wagner, A. *Nature Genet.* **24**, 355–361 (2000).
5. Taddei, F. *et al.* *Nature* **387**, 700–703 (1997).
6. Waddington, C. H. *Nature* **150**, 563–565 (1942).
7. McLaren, A. *Trends Genet.* **15**, 169–171 (1999).
8. Gibson, G. & Wagner, G. *Bioessays* **22**, 372–380 (2000).
9. Rutherford, S. L. & Lindquist, S. *Nature* **396**, 336–342 (1998).
10. Coyne, J. A., Barton, N. H. & Turelli, M. *Evolution* **51**, 643–671 (1997).
11. Metzgar, D. & Wills, C. *Cell* **101**, 581–584 (2000).

Geochemistry

Hot heads and cold tails

Nicholas Arndt

Every 30 million years or so, in short periods of violent activity, the Earth erupts enormous volumes of lava. Thick piles of flood basalt called large igneous provinces (LIPs) form either on the continents or in ocean basins. The biggest is the Ontong Java oceanic plateau, a pile of submarine lavas the size of a small continent. Many LIPs are related to currently active volcanoes. For example, the Etendeka-Paraná continental flood basalts of southern Africa and South America can be traced to active volcanoes on the small island of Tristan da Cunha in the southern Atlantic. The link between LIPs and modern volcanoes is explained by the theory of mantle plumes^{1,2}, which are upwellings of hot material from deep in the Earth. But the temperature of these plumes has long been debated. Thompson and Gibson³, on page 502 of this issue, report the discovery of unusually primitive olivine in volcanic rocks that raises important questions about the nature and evolution of mantle plumes.

Two influential models of mantle plumes^{4,5} propose that active 'hotspot' volcanoes, such as those of Tristan da Cunha, Iceland and Hawaii, form by melting in 'plume tails' — 100-km wide cylinders of upwelling hot mantle — whereas LIPs are attributed to melting in 'plume heads', which are enormous (500–2,000 km in diameter) spherical blobs of hot mantle on top of the plume tail. There are subtle differences

between these models in the way that the plume interacts with the surrounding mantle and the proposed source of the plume (Fig. 1).

The temperature within the mantle increases gradually with increasing depth until it hits a 'thermal boundary layer', where it jumps to a much hotter temperature. White and McKenzie⁴ propose that a plume head forms when a plume rising from a boundary layer between the lower mantle and the upper mantle (some 670 km deep) reaches the base of the cooler lithosphere (100–150 km below the surface). Campbell and Griffiths⁵ argue that the head grows and cools during ascent of the plume from the core-mantle boundary, a much deeper source 2,900 km below the surface. As a consequence, Campbell and Griffiths propose that the plume head is cooler than the tail, whereas in White and McKenzie's model the head could be hotter.

We have no direct means of testing these models. Using seismic tomography we can study the active plumes that feed Hawaiian or Icelandic volcanoes⁶, but the youngest major LIP, the Ethiopian flood basalts, formed 30 million years ago. To establish the geometry and temperature distribution within a LIP source we must resort to indirect methods. It is relatively straightforward to estimate a lava's temperature from its mineralogical or chemical composition. But complications arise because few erupted lavas are primary

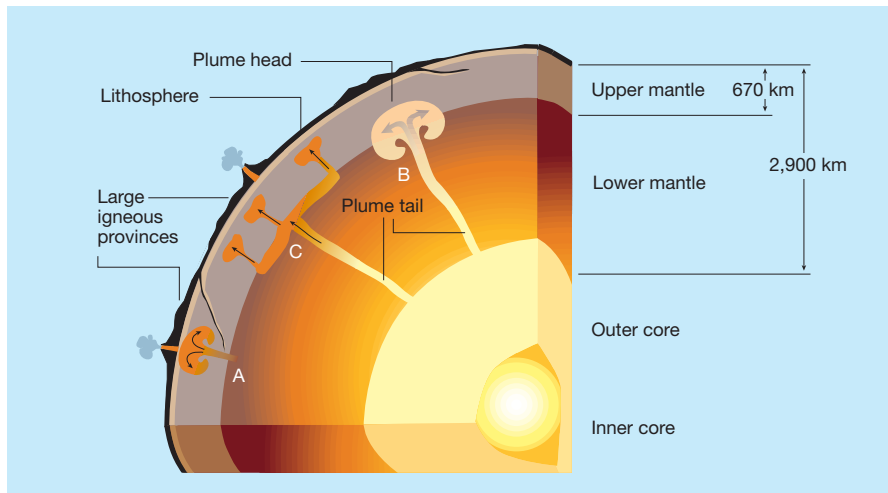


Figure 1 Different models for mantle plumes. These plumes are responsible for violent volcanic eruptions on the Earth's surface that leave behind thick piles of lava called large igneous provinces. In model A, a plume that rises from the boundary between the lower mantle and the upper mantle forms a head after it reaches the lithosphere⁴. In model B, a plume grows a large, and cooler, head as it ascends from a source at the core-mantle boundary⁵. Thompson and Gibson's discovery of primitive olivines in volcanic rocks would appear to support the first model, which allows for the plume head to be hotter than the tail. But model C, in which a plume stalls at the boundary between the lower mantle and the upper mantle and gives rise to smaller 'plumelets', offers a third explanation.

magmas — that is, direct melts of the mantle. Most partly crystallize to more 'evolved', less primitive compositions during ascent to the surface. To establish source temperatures, we must either seek out relatively rare magmas that have close-to-primary compositions, such as picrites and komatiites, or we must use relicts from primitive liquids trapped within more evolved lavas.

The latter approach was used by Thompson and Gibson³, who studied the Etendeka basalts, part of the South Atlantic LIP that was stranded on the African continent in present-day Namibia. They report the presence of unusually primitive olivine, containing up to 93.3% forsterite (Mg_2SiO_4), in basaltic dykes that form part of the Etendeka volcanic province. Most olivine crystals in these dykes have normal, less primitive compositions (84–90% forsterite). The forsterite content is important because it is related to the MgO content of the magma from which the olivine crystallized, which in turn can be used to estimate magma temperature.

Thompson and Gibson argue that the forsterite-rich olivine crystallized from a highly primitive (komatiite) parental magma that was trapped at the base of the crust. Some of these crystals were transported to the surface in magma of more normal, basaltic composition. The authors calculate that the komatiite magma contained about 24% MgO and formed from a plume head that was 300–400 °C hotter than the surrounding mantle, and significantly hotter than that of the sources of modern hotspot islands. They then argue that this conclusion supports models such as that of White and McKenzie, in which the plume head is hotter than the plume tail.

The significance of Thompson and Gibson's study is their discovery of highly primitive olivine and their conclusion that these crystals came from trapped magmas that were far hotter than the basalts that form most of the Etendeka province. The result supports studies showing that the sources of many LIPs are heterogeneous, probably with small regions of hot material and larger regions of cooler material. Komatiites and picrites in the 88 million-year-old Caribbean oceanic plateau provide direct evidence of highly primitive lavas^{7,8}, and primitive olivine is reported in other provinces⁹. Whether the Etendeka source was quite as hot as claimed is debatable, however, because of uncertainties in our knowledge of the influence of pressure, oxygen fugacity and the presence of water. So Thompson and Gibson's estimate could be up to 100 °C too high or too low.

I would also question whether the result necessarily supports the White and McKenzie model: a Campbell and Griffiths plume head is heterogeneous and contains material hot enough to generate the Etendeka picrites. In fact it may be a mistake to focus attention on only these two models. In more recent models, the form of the plume is quite different from that imagined by the pioneers of plume theory. It has been suggested, for example^{10,11}, that the plume does not penetrate the boundary between the upper mantle and the lower mantle, but spreads out as a sheet to give rise to a number of 'plumelets' (Fig. 1). To obtain definitive answers about the size, form and temperature distribution of the LIP source, the approach of Thompson and Gibson should be applied to other continental and oceanic plateaux. This, in



100 YEARS AGO

It is well known that while country-folk adhere to the old idea that adders when frightened are in the habit of protecting their young by swallowing them, a large number of naturalists regard the feat as an impossibility. In the September number of *The Zoologist* Mr. G. Leighton, a well qualified anatomist, has set himself the task of ascertaining whether there is any foundation for the objection. And he arrives at the conclusion that there is no anatomical reason why the oft-repeated statement of country observers should not be founded on fact. The author concludes by stating that the objections raised on the ground that the swallowing is unnecessary is a mere matter of opinion, adding that all that is now necessary is for a competent authority to dissect an adder which has been observed to swallow its young. 'Until this is done scientific naturalists will continue to regard the question as one capable of proof, if true, but hitherto unproved.'

From *Nature* 27 September 1900.

50 YEARS AGO

One session of the recent meeting at Birmingham of the British Association, held in the Great Hall of the University, was devoted to an account of industrial applications of atomic energy... It will not be possible to decide on the feasibility of nuclear power until pilot power-producers are built. Very rough figures of capital and fuel costs do no more than show that nuclear power costs will be likely to be of the same general order as conventional power costs... Design studies are being carried out on an experimental breeder reactor, one of which is already under construction in the United States. This experimental and development programme is likely to occupy the next decade. If the development work is successful, nuclear power would be likely to start in specialized applications where fuel costs are less important, and then to spread gradually to more conventional power stations. British fuel reserves have been estimated as lasting from 200 to 300 years; power consumption is increasing by 10 per cent a year and there is increasing competition with the chemical engineering industry for fuel. New power sources are therefore of considerable long-term importance, though other potentialities should be pursued in addition to nuclear power.

From *Nature* 30 September 1950.

combination with geochemical studies of erupted picrites and komatiites, would provide more solid information on conditions in the melting region. ■

Nicholas Arndt is at the Laboratoire de Géodynamique des Chaînes Alpines, Université Joseph Fourier, 1381 rue de la Piscine, 38041 Grenoble Cedex 9, France.
e-mail: arndt@ujf-grenoble.fr

1. Morgan, W. J. in *The Sea* (ed. Emiliani, C.) 443–487 (Wiley, New York, 1981).
2. Richards, M. A., Duncan, R. A. & Courtillot, V. E. *Science* **246**, 103–107 (1989).

3. Thompson, R. N. & Gibson, S. A. *Nature* **407**, 502–506 (2000).
4. White, R. & McKenzie, D. J. *Geophys. Res.* **94**, 7685–7729 (1989).
5. Campbell, I. H. & Griffiths, R. W. *Earth Planet. Sci. Lett.* **99**, 79–93 (1990).
6. Wolfe, C. J., Bjarnason, I. T., VanDecar, J. C. & Solomon, S. C. *Nature* **385**, 245–247 (1997).
7. Echeverria, L. M. *Contrib. Mineral. Petrol.* **73**, 253–266 (1980).
8. Echeverria, L. M. & Aitken, B. *Contrib. Mineral. Petrol.* **92**, 428–436 (1986).
9. Larsen, L. M. & Pedersen, A. K. J. *Petrol.* **41**, 1071–1098 (2000).
10. Thompson, P. F. & Tackley, P. J. *Geophys. Res. Lett.* **25**, 1999–2002 (1998).
11. Brunet, D. & Yuen, D. A. *Earth Planet. Sci. Lett.* **178**, 13–27 (2000).

Molecular physiology

Globins in the brain

Luc Moens and Sylvia Dewilde

The globins are proteins famed for their oxygen-carrying capacity. Two types are known in vertebrates: haemoglobin, found in the blood, and myoglobin, located in muscle. Now a third has been identified, as described by Burmester *et al.* on page 520 of this issue¹. It is found in the nervous system, mainly the brain, so the authors have called it 'neuroglobin'.

Haemoglobin is the protein responsible for the red colour of blood, and belongs to one of the best-studied protein families. The basic molecule, the globin chain, is a small globular protein with a relative molecular mass of ~17,000, carrying a porphyrin ring with a central iron atom. Haemoglobin-like molecules occur widely, in organisms ranging from bacteria to man. They transport and store oxygen, which is essential for the oxidative (aerobic) generation of energy by a cell's mitochondria; other functions include oxygen scavenging, transporting and detoxifying nitric oxide, and enzymatic activities². The other well-known globin, myoglobin, is involved in oxygen storage and facilitating oxygen diffusion to the mitochondria in muscle cells³.

Burmester *et al.* found neuroglobin, in both mice and humans, through database mining of genome-sequencing data. It is expressed at small but differential levels throughout the brain, and the authors suggest that it increases the oxygen availability to the tissue; this is the most likely explanation. The human neuroglobin gene maps to chromosome 14, and a comparative analysis of its protein product implies that members of the globin family diverged early in the evolution of multicellular animals.

The discovery of human neuroglobin is not completely surprising as neuroglobin-like proteins are known from the nervous systems of several invertebrate groups⁴. Their occurrence is not a general characteristic of these groups, however. These other neural globins probably occur because of

special needs created by the environmental conditions in which an organism lives. In such species, the functioning of the nervous tissue depends greatly on the oxygenation state of the associated globin. In *Aplysia*, a gastropod mollusc, for example, the firing activity of the neural ganglia (which effectively constitute the animal's central nervous system) is proportional to the degree of oxygenation of the neural globin⁵. And when *Tellina*, a bivalve mollusc, is kept in anaerobic conditions, neural excitability is sustained as long as oxygen can be delivered by the neural globin⁶.

Nerve tissue has a high energy demand: although the brain constitutes only about

2% of human body mass, it consumes about 20% of the available oxygen. The resulting energy is produced mainly by the aerobic combustion of glucose. It is used both to maintain ion gradients across cell membranes, which are essential for generating the action potentials necessary for nerve firing; and to power the motor proteins needed to transport materials throughout the nerve cell. A temporary, localized lack of oxygen (ischaemia) in brain tissue, as occurs in stroke, results in cell death and partial loss of nerve function. The oxygen stored by neuroglobin might help to maintain nerve function under these conditions, as it does in the invertebrate species. In this sense, neuroglobin might function in a similar way to myoglobin in cardiac muscle cells; as was confirmed last year in 'knock-out' mice lacking myoglobin⁷, myoglobin helps oxygen to diffuse to the mitochondria.

The data of Burmester and colleagues¹ suggest that neuroglobin expression in different regions of the human brain is inversely related to the sensitivity of these regions to ischaemic injury⁸. For example, the average duration of ischaemia producing half-maximal damage is 19.1 minutes in the cortex and 12.7 minutes in the hippocampus; expression of neuroglobin in the hippocampus is four times less than in the cortex⁹. One might speculate about a further clinical connection in diseases leading to neuronal degeneration. Lower resistance to ischaemia because of lower expression of neuroglobin, and thus reduced oxygen availability, could be a secondary factor in increasing cell death¹⁰. For

Conservation biology

Ducks and drakes

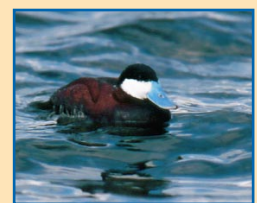
In Europe, the ruddy duck (*Oxyura jamaicensis*, right) was once seen as a welcome North American addition to wildfowl collections. But the species escaped, became widespread and has threatened the existence of the native white-headed duck, *O. leucocephalus*, by mating with the females. In Britain, the reluctant decision was taken to control ruddy ducks by shooting them.

But why should the bigger female white-headed ducks prefer the smaller ruddy-duck drakes rather than males of their own species? A clue comes from a paper by K. G. McCracken in *Auk* (**117**, 820–825; 2000). Studies of the related Argentine lake duck *O. vittata* reveal that the male has an especially large penis —

20 cm, when erect, which is about half of its owner's length. McCracken includes observations of drake ruddy ducks displaying their penises, apparently to impress the females; similar displays, where the males compete in groups for the favours of the females, occur in all the six other species of the *Oxyura* genus.

Whether penis size is an explanation for why female white-headed ducks preferentially mate with male ruddies remains a matter of speculation, however. Further anatomical and behavioural studies — a Kinsey report for ducks, as it were — are needed to resolve the question.

On the conservation front, the spread of ruddy ducks to Spain and Turkey is threatening



the white-headed populations there. In Spain, the interloper and any hybrids (which are fertile) are shot in an attempt to allow pure-breeding of *O. leucocephalus*. An encouraging report comes from one site near Alicante, on the Mediterranean coast. The estimated population in August this year was over 4,000, compared with 22 in the 1970s.

Chris Mead is at The Nunnery, Hilborough, Thetford, Norfolk IP26 5BW, UK.
e-mail: chris.mead@zetnet.co.uk

example, it is striking that in Alzheimer's disease the hippocampal region — which has the lowest expression of neuroglobin — is frequently affected by neurofibrillary tangles and senile plaques, which are the morphological hallmarks of the disease characterized by neuronal degeneration¹¹.

The discovery of neuroglobin in mice and humans, and the sporadic occurrence of neural globins in invertebrates, suggests that these molecules probably evolved with the evolutionary development of the nervous system itself, and that they will be found in all taxonomic groups. Beyond that, neuroglobin's likely involvement in supplying oxygen to nerve cells may lead to the possibility of improving neuron function and survival in both health and disease. ■

Luc Moens and Sylvia Dewilde are in the Department of Biochemistry, University of Antwerp, Universiteitsplein 1, B-2610 Wilrijk, Belgium.
e-mail: lmoens@uia.ua.ac.be

1. Burmester, T., Weich, B., Reinhardt, S. & Hankeln, T. *Nature* **407**, 520–523 (2000).
2. Minning, D. M. *et al. Nature* **401**, 497–502 (1999).
3. Wittenberg, B. A. & Wittenberg, J. B. *Proc. Natl Acad. Sci. USA* **84**, 7503–7507 (1987).
4. Vandergon, T. L. *et al. J. Biol. Chem.* **273**, 16998–17011 (1998).
5. Chalazonitis, N. *et al. C. R. Seances Soc. Biol. Fil.* **159**, 240–244 (1965).
6. Kraus, D. W. & Colacino, J. M. *Science* **232**, 90–92 (1986).
7. Gödecke, A. *et al. Proc. Natl Acad. Sci. USA* **96**, 10495–10500 (1999).
8. Adams, J. H., Corsellis, J. A. N. & Duchen, L. W. *Greenfield's Neuropathology* (Butler & Tanner, London, 1984).
9. Zivin, J. A. *Neurology* **50**, 599–603 (1998).
10. de la Torre, J. C. *Ann. NY Acad. Sci.* **826**, 75–91 (1997).
11. Hendrick, L. & Van Broeckhoven, C. *Eur. J. Biochem.* **237**, 6–15 (1996).

Optical astronomy

The resolution revolution

Tyler Nordgren

Astronomers are fond of 'standard candles' — objects, such as supernovae or Cepheid variable stars, whose brightness is related in a predictable way to their true distance from us. On page 485 of this issue, Lane *et al.*¹ report the first unambiguous detection of the pulsating diameter of a Cepheid variable star. The observations were made using the Palomar Testbed Interferometer and exploit the continuing revolution in telescope resolution to attempt to answer one of the hottest questions in extragalactic astronomy: how big is the Universe?

Stars make up the vast majority of objects visible to the naked eye yet the average star looks the same to the casual star gazer as it does through the largest telescopes on Earth. Even NASA's Hubble Space Telescope sees all but the largest star in the sky, Betelgeuse, as unresolved points of light² (Fig. 1a). As the diameter of a telescope's mirror increases, so does the angular resolution (the ability to discern fine detail). But the blurring effect of the Earth's atmosphere means that the finest resolution available to ground-based optical telescopes is only about 0.8 arcseconds (1 arcsecond is 1/3,600 of a degree). The giant star Betelgeuse is only 50 milliarcseconds in diameter, whereas most stars visible to the naked eye are ten to a hundred times smaller. With recent advances in adaptive optics this blurring effect can be largely removed from ground-based telescopes, yet engineering and cost constraints still limit the size of a single mirror and thus the maximum possible angular resolution.

An interferometer combines the light from two or more widely spaced mirrors, resulting in an array with the same resolution as a single large telescope of the same diam-

eter as the separation between the mirrors. So radio astronomers using the Very Long Baseline Array are able to simulate a single radio telescope the size of North America. Unfortunately, most stars emit radiation at optical rather than radio wavelengths, and technical obstacles have prevented optical telescopes being built on this scale. The Palomar Testbed Interferometer (PTI), located on Palomar Mountain in California, has the resolving capability of a single 110-metre infrared telescope — much larger than any single telescope built so far, and good enough to resolve stars as small as a few milliarcseconds in diameter³.

Optical interferometers are now operating or under construction in Britain, continental Europe, Australia, Hawaii and North and South America. In the past decade, interferometers have successfully measured stellar angular diameters smaller than two milliarcseconds with precision greater than 10% (ref. 4). With the increase in size and resolution of these instruments has come an increase in the number of fields in which contributions and discoveries can be made. One such area is the size and age of the Universe.

Cepheid variables are supergiant stars that regularly change in size and brightness. Numerous Cepheids have been found in our own and other galaxies including the Large Magellanic Cloud (LMC), a satellite galaxy of the Milky Way. Observations of the Cepheids in the LMC, all of which are about the same distance away, reveal a direct relation between pulsation period and intrinsic brightness (luminosity). Measure the period of a Cepheid's pulse cycle, and the derived luminosity combined with its apparent brightness yield its distance. Identify Cepheids in a nearby galaxy and one immediately knows that galaxy's distance (a key project of the Hubble Space Telescope). These distances are used to help derive the Hubble constant — a measure of the expansion rate of the Universe — which is the most direct indicator of the size and age of the Universe⁵.

For this method to work, we also need an independent measure of the distance to the Cepheids that allowed us to derive the period–luminosity relation in the first place. With the size and age of the Universe at stake, the distance to the LMC is one of the most hotly debated values in astronomy⁶. Finding a simple way to measure distances to enough Cepheids in the LMC would allow one to calibrate their overall period–luminosity

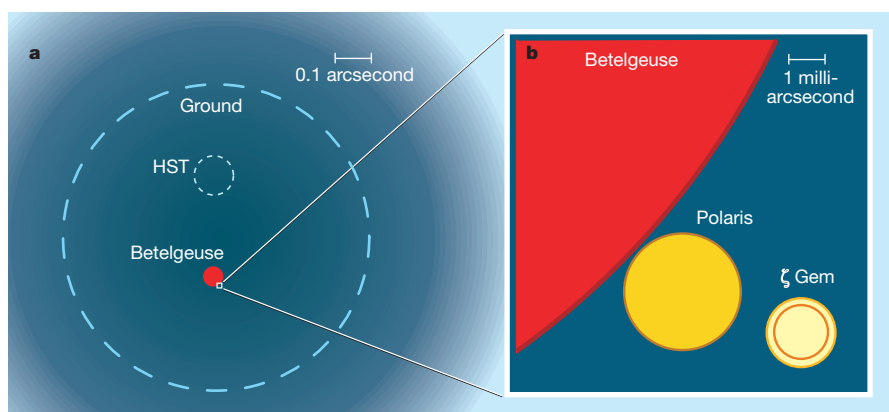


Figure 1 Getting up close and personal to a star. a, The relative sizes of the average resolution available to a ground-based telescope because of atmospheric interference and the diffraction-limited resolution of the Hubble Space Telescope (blue dashed circles). The scale bar represents 0.1 arcseconds, where 1 arcsecond is 1/3,600 of a degree. The angular size of the largest star in the sky, the red supergiant Betelgeuse, is shown for comparison. b, An enlargement of a box 10 milliarcseconds in size, showing part of Betelgeuse and, for comparison, the warmer supergiant Polaris, the North Star. The minimum angular diameter (red contour) of the pulsating Cepheid variable ζ Gem is only 10% smaller than its maximum angular size. For comparison, a person standing on the Moon is only 1 milliarcsecond in height as seen from the Earth.

Translation

The twenty-first amino acid

John F. Atkins and Raymond F. Gesteland

relation. One such way is to measure the change in the Cepheid's angular diameter as it pulsates. A star's angular diameter is a combination of its true diameter and its distance from the Earth. A change in its true diameter can easily be obtained from traditional spectroscopy. Spectroscopic observations record the radial velocity of the Cepheid atmosphere as it pulsates, and integrating these velocities determines the change in the radius of the star. The ratio of the change in the star's true radius and the change in its angular diameter gives the distance.

Zeta Geminorum (or ζ Gem, one of the four brightest Cepheid variable stars in the Northern Hemisphere) has an estimated angular diameter of around 1.5 milliarcseconds with an expected diameter variation of 10% (Fig. 1b). Whereas interferometric observations of all four Cepheids have been made previously using other interferometers, none were able to unambiguously detect diameter variations⁷. With the results presented here, Lane *et al.*¹ can claim to have the first conclusive measurements of a Cepheid's changing diameter. Furthermore, the distance derived from these observations is in perfect agreement with the rough distance to ζ Gem already obtained by the European Space Agency's Hipparcos satellite⁸. The PTI observations provide a new and simple method for determining the distance to more Cepheids at larger distances.

Although the first optical interferometer was built in the 1920s (ref. 9), astronomers have had to wait for advances in computer and optics technology before the widespread development of optical interferometers such as PTI became possible. Because these are complex machines, interferometry researchers have invariably had to spend considerable time convincing the astronomical community of the accuracy of their results⁴. In their observations of ζ Gem, Lane *et al.*¹ have done an excellent job of comparing their observations to a wide range of previous Cepheid observations. With new-found confidence in these difficult observations, optical interferometry is now ready to move beyond confirming previous discoveries to making new ones of its own.

Tyler Nordgren is at the US Naval Observatory Flagstaff Station, PO Box 1149, Flagstaff, Arizona 86001, USA.

e-mail: nordgren@no.navy.mil

1. Lane, B. F., Kuchner, M. J., Boden, A. F., Creech-Eakman, M. & Kulkarni, S. R. *Nature* **407**, 485–487 (2000).
2. Gilliland, R. L. & Dupree, A. K. *Astrophys. J. Lett.* **463**, L29–L32 (1996).
3. van Belle, G. *et al. Astron. J.* **117**, 521–533 (1999).
4. Nordgren, T. E. *et al. Astron. J.* **118**, 3032–3038 (1999).
5. Mould, J. R. *et al. Astrophys. J.* **529**, 786–794 (2000).
6. Nelson, C. A., Cook, K. H., Popowski, P. & Alves, D. R. *Astron. J.* **119**, 1205–1213 (2000).
7. Mourard, D. *et al. Astron. Astrophys.* **317**, 789–792 (1997).
8. Feast, M. W. & Catchpole, R. *Mon. Not. R. Astron. Soc.* **286**, L1–L5 (1997).
9. Michelson, A. A. & Pease, F. G. *Astrophys. J.* **53**, 249–259 (1921).

In the early days of molecular biology, it was thought that only 20 amino acids are specified by the genetic code. Four 'letters', or nucleotides, are used in this code, which is read by grouping the letters into triplets. When the genetic code was cracked in the 1960s, each of the 64 possible combinations of three letters was thought to have just one use, either encoding one of the 20 amino acids or marking the start or end point of the read-out. But in 1986 it was discovered that under special circumstances the meaning of one of the 'stop codons' — the letters UGA — is changed to specify the twenty-first amino acid, selenocysteine. Only a few proteins have selenocysteine in their sequences, but this unusual amino acid often has a very specific chemical role. The mechanisms for encoding selenocysteine are proving very interesting. Tujebajeva and co-workers¹ and Fagegaltier and colleagues², writing in *EMBO Reports* and *EMBO Journal*, respectively, provide clues to how mammalian cells read 'selenocysteine' instead of 'stop'.

First, let us remind ourselves of the background. To translate a gene into a protein, a cell first makes a working copy of the gene — a messenger RNA template. RNA-protein complexes called ribosomes then read along the template, deciphering what each nucleotide triplet means. They start at the triplet AUG, adding amino acids sequentially (brought in by transfer RNAs) as specified by the following triplets in the

sequence, and stop when they reach a stop codon. Unless, that is, the stop codon is UGA and the mRNA template contains instructions that a selenocysteine residue should be inserted at that point.

The mechanism by which UGA is redefined to specify selenocysteine in the bacterium *Escherichia coli* was unravelled by Böck and colleagues³. Here, the key instruction is a special folded structure in the mRNA, a 'stem loop', immediately following the UGA. The loop of this stem loop binds a complex consisting of a protein, SelB, and the special transfer RNA that carries selenocysteine. SelB is an 'elongation factor', like the better-known EF-Tu, which brings to ribosomes the tRNAs that carry the standard amino acids. SelB, by contrast, is specific for the selenocysteine-bound tRNA. It also has a feature not seen in EF-Tu: a carboxy-terminal extension that is responsible for binding to the loop sequence of the stem loop, ensuring that only a UGA codon followed by the right stem-loop structure will have its meaning changed. So, the proximity of the UGA codon and the stem loop leads to a model in which the selenocysteine-bound tRNA is delivered, by SelB, directly to the UGA codon.

But in mammals the signal that redefines UGA is a more complicated stem-loop structure (called SECIS) that is located not next to the UGA codon, but far away, beyond the end of the coding sequence of the mRNA⁴

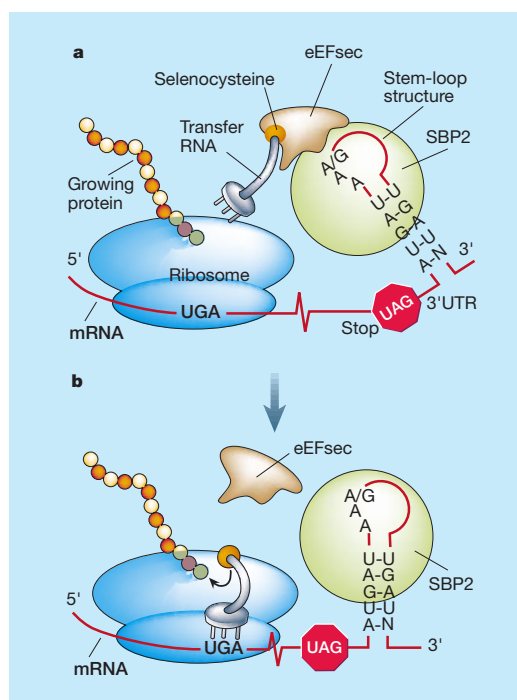


Figure 1 Redefining the stop codon in mammalian messenger RNAs. Ribosomes move along mRNA, deciphering the nucleotide sequence and making a protein according to the encoded amino-acid sequence. The nucleotide sequence UGA normally specifies that the ribosome should stop translation. But sometimes this stop codon can be redefined, so that the twenty-first amino acid — selenocysteine — is incorporated instead. This model shows how this might be done¹¹. a, A 'stem loop' structure in the downstream, untranslated part of the mRNA binds to a protein called SBP2. SBP2 in turn binds to the eEFsec protein, which itself has recruited the transfer RNA carrying selenocysteine. b, The selenocysteine-bound tRNA is then delivered to the waiting UGA, for incorporation into the growing amino-acid string that constitutes the newly created protein.

(Fig. 1). Attempts to understand how this system works focused first on trying to find a protein that binds to the SECIS region, on the assumption that such a protein might work in a similar way to the bacterial SelB protein. Nevertheless, despite numerous false alarms, the mammalian counterpart(s) of the bacterial SelB remained elusive. These 'scarlet pimpernel' properties can be explained by the discovery that, in mammals, the functions of SelB are divided into two proteins.

One protein, called SBP2 (SECIS-binding protein 2), binds the SECIS element⁵. The binding specificity⁵ of this protein depends on a key feature of the SECIS — a quartet of non-Watson–Crick base pairs⁶. But this protein does not have the task of binding to the selenocysteine-carrying tRNA, and it does not have the sequence features that would be expected of a protein that brings tRNAs to the ribosome. The discovery of the mammalian protein that does this — called eEFsec^{1,2} — was helped by the finding⁷ of a specialized elongation factor in *Methanococcus jannaschii*, a microorganism from the Archaea. This elongation factor from *M. jannaschii* does not bind to the SECIS element, but does bind to the selenocysteine tRNA.

Tujebajeva *et al.*¹ and Fagegaltier *et al.*² have now discovered the mammalian counterpart of this archaeal protein, by searching through sequence databases using the amino-acid sequence of the archaeal protein as a starting point. Like its archaeal counterpart, the mammalian protein does not bind the SECIS element but does interact directly with both tRNAs bearing selenocysteine^{1,2} and SBP2 (ref. 1). So the SECIS element — through a two-protein complex containing SBP2 and eEFsec — can recruit selenocysteine-carrying tRNAs (Fig. 1).

But this protein complex, when bound to the stem-loop structure, is a long way from the UGA codon. How does the distant complex find the waiting ribosome and deliver selenocysteine? This is especially perplexing in the case of a protein called SelP, whose mRNA has between 10 and 17 UGA codons depending on the species^{8,9}, each coding for selenocysteine. For such cases, one could imagine a processive model based on the known proximity of the two ends of an mRNA strand. In this model the SECIS element would deliver the SBP2 complex to a ribosome that is just starting translation. When the ribosome reaches the first UGA codon, the required selenocysteine is already with the ribosome, ready to be inserted into the growing protein. Afterwards, the ribosome-bound tRNA might be able to pick up a second tRNA-bound selenocysteine, and so on. But, if operative, this or other processivity models must have sophisticated aspects that are not yet apparent¹⁰.

The obvious model, by analogy with the

situation in *E. coli*, is that the protein complex, bound to the tRNA and to the SECIS element, reaches back and delivers the tRNA (plus selenocysteine) directly to the ribosome¹¹ (Fig. 1). But how the loaded SECIS finds the waiting ribosome remains a mystery. With the new proteins^{1,2,5} now identified, one hopes that answers to these questions will not be long in coming. Help will no doubt come from studies of the supramolecular translation complexes¹² that may coordinate interactions central to the decoding of genetic text.

John F. Atkins and Raymond F. Gesteland are in the Department of Human Genetics, University of Utah, Salt Lake City, Utah 84112-5330, USA.

Statistical physics

Following the crowd

David J. Low

When we are at a major sporting event or travelling on public transport, our safety and comfort depend crucially on our fellow crowd members and on the design and operation of the facility we are in. So it is unnerving to realize that the modelling currently used to design and operate these venues has more in common with the design of water-pipe networks than anything with a human dimension. On page 487 of this issue¹, Helbing, Farkas and Vicsek present a dramatically different and potentially far more realistic modelling approach.

The movement of large numbers of people is important in many situations, such as the evacuation of a building in an emergency. In large crowds there is a risk of injury,

e-mails: atkins@howard.genetics.utah.edu
ray.gesteland@genetics.utah.edu

1. Tujebajeva, R. M. *et al.* *EMBO Rep.* **1**, 158–163 (2000).
2. Fagegaltier, D. *et al.* *EMBO J.* **19**, 4796–4805 (2000).
3. Böck, A. *BioFactors* **11**, 77–78 (2000).
4. Berry, M. J. *et al.* *Nature* **353**, 273–276 (1991).
5. Copeland, P. R., Fletcher, J. E., Carlson, B. A., Hatfield, D. L. & Driscoll, D. M. *EMBO J.* **19**, 306–314 (2000).
6. Walczak, R., Carbon, P. & Krol, A. *RNA* **4**, 74–84 (1998).
7. Rother, M., Wilting, S., Commans, S. & Böck, A. *J. Mol. Biol.* **299**, 351–358 (2000).
8. Hill, K. E., Lloyd, S. & Burk, R. F. *Proc. Natl Acad. Sci. USA* **90**, 537–541 (1993).
9. Tujebajeva, R. M., Ransom, D. G., Harney, J. W. & Berry, M. J. *Genes Cells* (in press).
10. Nasim, M. T. *et al.* *J. Biol. Chem.* **275**, 14846–14852 (2000).
11. Low, S. C. & Berry, M. J. *Trends Biochem. Sci.* **21**, 203–208 (1996).
12. Stapulionis, R., Kolli, S. & Deutscher, M. P. *J. Biol. Chem.* **272**, 24980–24986 (1997).

and even loss of life, owing to the enormous forces that can be exerted on a single individual by the surrounding throng. The pressures that build up can bend steel barriers or push down brick walls. The consequences of crushing, trampling and panic in crowds are well known^{2,3}, and a proper understanding of how groups of people move is vital if we are to minimize risks in these situations.

The traditional approach to predicting the motion of large crowds of pedestrians models the crowd as if it were a continuous homogeneous mass that behaves like a fluid flowing along corridors. The predicted flow rate indicates how quickly the crowd will move. However, this traditional approach assumes that the crowd is made up of identi-

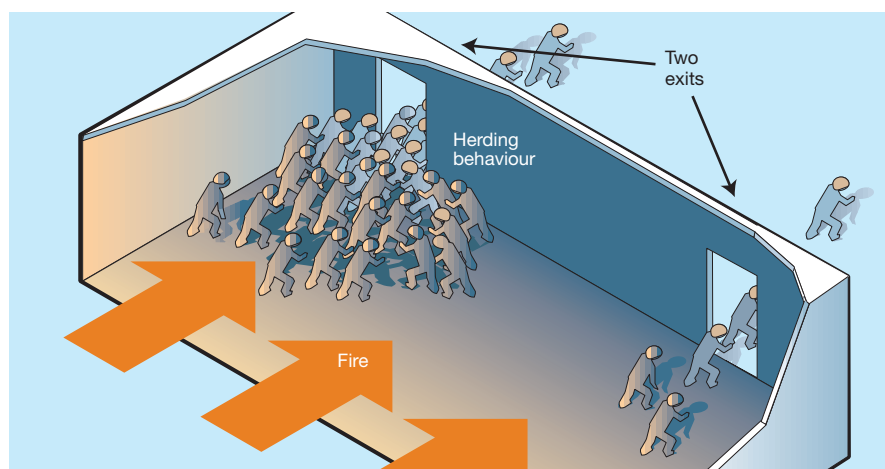


Figure 1 How crowd behaviour affects escape from a smoke-filled room. Previous simulations of pedestrian behaviour in crowds have used a model based on fluid flow through pipes, but these ignored the actions of individuals. According to the individual-centred model of Helbing *et al.*¹, the evacuation of pedestrians from a smoke-filled room with two exits can lead to herding behaviour and clogging at one of the exits. By contrast, a traditional fluid-flow model would predict the efficient use of both exits. A more individual-centred approach is required to reproduce the behaviour of real crowds.

cal, unthinking elements. A fluid particle cannot experience fear or pain, cannot have a preferred direction of motion, cannot make decisions, and cannot stumble or fall. The diverse behaviour of individual crowd members can drastically change the way in which the crowd as a whole behaves. Injuries to crowd members are related less to average pressure within the crowd than to point pressures at individual localities, and an injured pedestrian can fall and become an obstacle to the movement of others.

The new approach requires a recognition that the crowd is made up of individuals who possess the ability to think and react to events around them. One of the most dramatic cases in which human behaviour influences events is in a panic situation. Helbing *et al.* address this particular problem with a computer model of pedestrian behaviour that includes realistic reactions to crushing, panic and reduced visibility. They also simulate the tendency of people to do what others do and 'follow the crowd', but also allow for individuals to adopt personal strategies. Helbing *et al.* demonstrate that, because of their increased speed, panicking individuals will block up an exit that they could pass through safely at normal walking speed. They also show that a widening in a corridor actually slows down the movement of pedestrians, rather than allowing them to move faster, as one would assume. This surprising result is explained by those pedestrians who might have tried to move away from or overtake each other having to squeeze back into the mainstream flow at the end of the widening.

Modelling a crowd composed of discrete individuals rather than a continuous fluid clearly brings added complications. Helbing *et al.* model 'non-fluid' crowd properties, such as the 'faster-is-slower' phenomenon in which people in a rush end up going slower. They also investigate the best evacuation strategy for people in a smoke-filled room (Fig. 1). Such information can then be used to work out low-risk designs for the width of corridors, the number and position of doors, and the size of areas where people may gather. But these types of study can also provide us with a wider range of possible solutions to crowd problems. The crowd composed of individual people can respond to information directed towards them, to help them choose the most appropriate direction to take or the most appropriate exit to use.

In the past, one of the main barriers to adopting this approach was the enormous number of calculations that are required to solve separate equations of motion for each crowd member. Modern computing power has dramatically changed that situation. Indeed, individual-centred approaches are now widely used in the modelling of road traffic networks^{4,5}, which also used to be dominated by fluid-flow models. Similar

individual-centred traffic models⁶ have produced excellent results and have led to effective new traffic management strategies. There has been extremely strong financial motivation to produce such improved traffic models. Traffic management strategies and road-building projects cost enormous sums of money, as do the delays caused by road congestion, and improved traffic modelling techniques can produce considerable savings. But the potential benefit of improved pedestrian models is even more valuable — a reduction in personal injury.

The model presented by Helbing *et al.* is just one of many possible models. To decide whether a particular model is an accurate description of real life, or to determine which model is the 'best' for the situation under consideration, requires real data to compare with each model's predictions. But such data are scarce or non-existent and may be extremely difficult to collect. With any type of mathematical modelling we always have to

be careful to distinguish between 'real life' and our attempt to model it. Failing to recognize this difference can have serious consequences. But provided we are aware of when it is appropriate to use a particular model, it can provide valuable information to guide the planning process, for construction and for dealing with emergencies. Perhaps perfect safety is unattainable, but improved models of crowd dynamics can help to increase our safety in crowded situations. ■

David J. Low is in the Department of Civil and Offshore Engineering, Heriot-Watt University, Riccarton Campus, Edinburgh EH14 4AS, UK. e-mail: d.j.low@hw.ac.uk

1. Helbing, D., Farkas, I. & Vicsek, T. *Nature* **407**, 487–490 (2000).
2. Taylor, P. *The Hillsborough Stadium Disaster*, 15 April 1989. Final Report (Home Office, HMSO, London, 1990).
3. Johnson, N. R. *Social Prob.* **34**, 362–373 (1987).
4. Addison, P. S. & Low, D. J. *Chaos* **8**, 791–799 (1998).
5. Low, D. J. & Addison, P. S. *Nonlin. Dynam.* **16**, 127–151 (1998).
6. Duncan, G. *Paramics-MP Final Report* (Edinburgh Parallel Computing Centre, Univ. Edinburgh, 1994).

Genomics

Use your neighbour's genes

Don Cowan

On page 508 of this issue, Ruepp and colleagues¹ describe the complete genome sequence of the acid- and heat-loving microorganism *Thermoplasma acidophilum*. This hardy organism, which lacks a cell wall, grows best on organic substrates at pH 2 and 59 °C. It was first isolated, in the late 1960s, from a self-heating ore pile²; such ore piles generate heat through internal microbial activity.

Microbial physiologists and structural biologists have long been fascinated by the ability of this microorganism to grow at high temperatures and low pH without the structural protection of a conventional cell wall. *T. acidophilum* is also interesting from an evolutionary perspective. Its cellular morphology seems primitive, and it contains complexes involved in protein folding, degradation and turnover that look like simple versions of related structures in eukaryotic cells (loosely, those cells with a nucleus — the type of cell that makes up higher organisms such as you and me). These facts intrigue evolutionary biologists, who have speculated that *T. acidophilum* is an ancestor of the eukaryotic cell.

Initially, *T. acidophilum* was classified as a thermophilic mycoplasma — a heat-loving example of a group of primitive, gliding bacteria, which lack cell walls². But following analysis of its lipid composition and ribosomal RNA sequences, it was reassigned to the new 'third domain of life' — the Archaea³ (Fig. 1). *T. acidophilum* is the ninth member

of the Archaea for which the genome has been completely sequenced^{4–10}. All except one of these microorganisms are heat-loving. Why is there this focus on the ther-

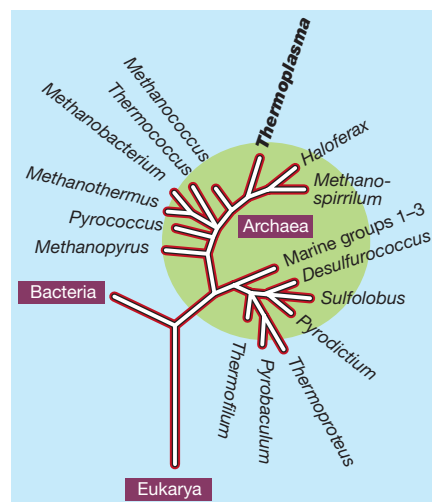


Figure 1 The three domains of life, showing how *Thermoplasma acidophilum* — the latest archaeon whose genome has been sequenced¹ — fits into the evolutionary scheme of things. On the basis of its primitive morphology and the presence of what seem to be several primitive cellular structures, *T. acidophilum* was once thought to be an ancestor of eukaryotic cells. But details of its genome sequence¹ make this unlikely. This phylogenetic tree is based on 16S ribosomal RNA sequences (modified from ref. 12).

mophilic Archaea? The explanation is three-fold and lies in the intense interest in the biochemistry of these unusual organisms, in the possibility that they represent the earliest forms of life, and in the biotechnological potential of their genes and gene products.

Thermoplasma acidophilum has one of the smallest of the archaeal genomes to have been sequenced so far. Even so, the speed at which Ruepp *et al.*¹ sequenced the genome was remarkable. The full sequence of over 1.5 million base pairs was obtained from only 7,855 sequencing reactions — an effective yield of 199 base pairs per reaction, compared with the 66 base pairs per reaction for the slightly larger genome of the hyperthermophilic bacterium *Thermotoga maritima*¹¹. The keys to this efficiency were the use of sequencing vectors containing very large DNA inserts; an extended-sequencing method referred to as 'primer walking'; and a policy of stopping the sequencing of an insert when the primer walking encountered a stretch of DNA whose sequence was already known.

So, what can we learn from this genome? The most startling observation is the high proportion of genes that seem to have been acquired from other species. For example, 17% of all identified 'open reading frames' (the parts of genes that encode proteins) have relatives in the not-yet-completely sequenced genome of the archaeon *Sulfolobus solfataricus*.

There may be several reasons why *T. acidophilum* has such an extraordinary ability to acquire external genes. First, environmental proximity is clearly important. Microorganisms of the genus *Sulfolobus* might be the most common archaeal species in the habitat occupied by *T. acidophilum*. In addition, a further 17% of the open-reading frames of *T. acidophilum* are 'bacteria-like'. So it might also have acquired some of its genes from bacteria such as *Alicyclobacillus*, *Thiobacillus* or *Sulfobacillus*, the habitats of which overlap with that of *Thermoplasma*. Second, the absence of a conventional, protective cell wall could be particularly significant: a cell wall is a major barrier to the entry of large molecules into a cell. Finally, the *T. acidophilum* genome might not be protected by a restriction/modification system, a set of enzymes designed to recognize and destroy foreign DNA. The organism has no 'restriction endonuclease' activity, although its genome might encode a DNA methyltransferase, normally part of a restriction/modification system, and restriction endonuclease genes have little sequence similarity and cannot be recognized in a gene sequence.

Interestingly, the *Sulfolobus*-like genes in the *T. acidophilum* genome are clustered into several (at least five) discrete regions. Ruepp *et al.* conclude that only a few gene-transfer events occurred, each involving movements of large chunks of genetic sequence. But the

transfer of smaller gene fragments between species tends to be more common, raising the question of why this seems not to have happened for *T. acidophilum*.

One issue can be almost settled by the details of this new genome sequence: whether *T. acidophilum* is an ancestor of eukaryotic cells. Ruepp *et al.* compared *T. acidophilum* genes with those in bacterial and eukaryotic databases. The results show that, if anything, the *T. acidophilum* genes are more similar to bacterial genes than to eukaryotic ones. Key 'marker' genes found in eukaryotes (such as genes encoding subunits of the nuclear pore complex) are not found in the *T. acidophilum* genome.

Finally, on a different note, the completion of another genome sequence reminds us how much we still do not know about gene function as a whole. Of the predicted 1,509 open reading frames in the *T. acidophilum* genome, 29% are akin only to 'hypothetical' open reading frames in other genomes, and 16% have no relatives elsewhere. This means that, as yet, we do not know what 45% of the protein-coding regions in the *T. acidophilum* genome do. That is a lot of genes. These percentages are typical for newly sequenced genomes. But the results serve as a reminder of the need both for more advanced data-mining techniques (which would increase our ability to pick out similar sequences from different genomes and to identify putative functions) and for the continuation of more classical molecular and functional research. ■

Don Cowan is in the Department of Biochemistry and Molecular Biology, University College London, Gower Street, London WC1E 6BT, UK.

e-mail: don.cowan@ucl.ac.uk

1. Ruepp, A. *et al.* *Nature* **407**, 508–513 (2000).
2. Darland, G. *et al.* *Science* **170**, 1416–1418 (1970).
3. Woese, C. R. & Fox, G. E. *Proc. NY Acad. Sci.* **74**, 5088–5090 (1977).
4. Kawarabayasi, Y. *et al.* *DNA Res.* **6**, 83–101 (1999).
5. Klenk, H. P. *et al.* *Nature* **390**, 364–370 (1997).
6. Smith, D. R. *et al.* *J. Bacteriol.* **179**, 7135–7155 (1997).
7. Bult, C. J. *et al.* *Science* **273**, 1058–1073 (1996).
8. Maeder, D. L. *et al.* *Genetics* **152**, 1299–1305 (1999).
9. Kawarabayasi, Y. *et al.* *DNA Res.* **5**, 55–76 (1998).
10. Kawashima, T. *et al.* *Proc. J. Acad. B* **75**, 213–218 (1999).
11. Nelson, K. E. *et al.* *Nature* **399**, 323–329 (1999).
12. Pace, N. R. *Science* **276**, 734–739 (1997).

Corrections

In the News & Views article "Global change: Plankton cooled a greenhouse" by Birger Schmitz (*Nature* **407**, 143–144; 2000), the period of 'superwarm' conditions at the Palaeocene/Eocene boundary should have been cited as lasting about 60,000 (not 150,000) years. Proof corrections made to the paper concerned (by S. Bains *et al.* **407**, 171–174; 2000) were not passed on to the News & Views author.

In Jim Gillon's "Earth systems: Feedback on Gaia" (*Nature* **406**, 685–686; 2000) a huge expansion of vegetation was cited as having taken place 550 million years ago. The generally accepted date for that expansion is around 400 million years ago.

Daedalus

Talking to animals

Last week Daedalus presented his mobile 'Ultraphone' for silent speech. The user whispers or mouths his message silently. His voiced tone is replaced by an inaudible ultrasonic tone launched into his mouth. His tongue and palate modulate this into inaudible ultrasonic speech. A heterodyne circuit downshifts it back to an audio signal, which is transmitted.

Daedalus now has a biological use for the technique. Many animals — cats, dogs, bats, rats and many insects — communicate partly or mainly in the ultrasonic band, above human hearing. Instruments can downshift their voices into the audio band, as the Ultraphone does. But the Ultraphone also performs the converse operation. It upshifts the human voice into a range much more significant to animal ears.

So DREADCO animal keepers are now speaking to their charges on an adapted Ultraphone. Its intense ultrasonic beam lets them 'speak' loudly and clearly in this waveband. Its wideband microphone and heterodyne circuit downshifts the animals' ultrasonic responses back into the audio for human recognition. The ultrasonic frequency of the 'Animal Ultraphone' can be adjusted to the band in which the conversation seems to flow most freely.

Laboratory rats are the first subjects. These bright and companionable rodents can be tamed quite well by normal means. An added vocal channel should bring them even closer to human understanding. The DREADCO staff are trying to imitate their language, and give them ultrasonic 'names' that they can recognize. Bats, those ultrasonic experts, may also respond to the Ultraphone; insects are probably too dim.

But the main impact of the Animal Ultraphone will be on domestic cats. These intelligent pets are famous for their aloof independence. They notoriously refuse to learn (or at any rate to obey) even the simplest commands. To them, human speech is probably a low vague mumbling. But talked to in the right ultrasonic band, even the most suspicious cat should become alert and responsive. It might even reply in tones that, downshifted to mellow audio by the Animal Ultraphone, appeal to the human ear.

Feline-human relations will be transformed. Bonded at last by mutually appreciated vocal expression, cats and cat-lovers will bestow far more warmth and company upon each other, bringing a new and welcome closeness to a long-standing association. And the true intelligence of these friends of Man may at last become apparent.

David Jones

Obituary

Mark Oliphant (1901–2000)

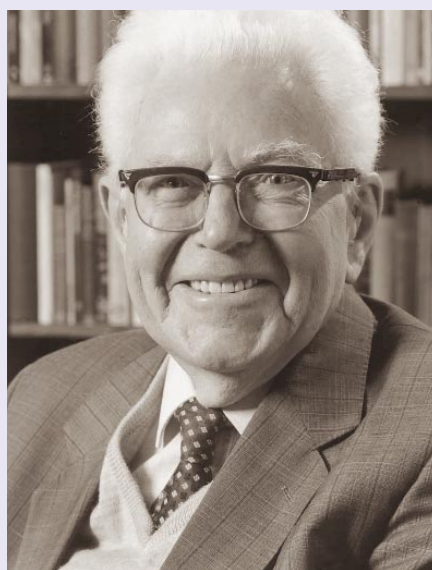
The Australian physicist Sir Marcus Laurence Elwin Oliphant, who died on 14 July 2000, was the last of the 'Rutherford Boys' — the brilliant team who, under the leadership of Lord Rutherford, made the Cavendish Laboratory in Cambridge the Mecca of nuclear research during the years between the two world wars. A figure larger than life, Oliphant exuded physical vitality and self-confidence. A genial extrovert, with a booming laugh, he spoke his mind bluntly, never shy of challenging authority. As a scientist he was imaginative and creative; as a public figure he had a strong commitment to freedom of speech and an abhorrence of secrecy in any walk of life.

Oliphant joined the Cavendish Laboratory in 1927, on an 1851 Exhibition Scholarship. He quickly endeared himself to Rutherford by his experimental skill. Making things with one's own hands was in the Cavendish tradition of 'sealing-wax and string', but Oliphant was also an enthusiast for new technologies, with a penchant for building particle accelerators.

Improving on the Cockcroft–Walton accelerator by designing ion sources of much greater intensity, Oliphant's main contributions were the study of the nuclear reactions that occurred when deuterium was bombarded with deuterons, which led to the discovery of tritium. Considering his later anti-nuclear campaigns, it is ironic that the main practical applications of his work should be in nuclear weapons — tritium is a booster for the fission bomb, and the fusion reactions are the basis of the hydrogen bomb.

After moving to Birmingham in 1937, Oliphant took on the task of converting a moribund physics department into a leading nuclear research centre. He saw to this by building a 60-inch cyclotron, the largest at the time in England. However, the completion of the machine was interrupted by the outbreak of the Second World War in 1939. Two projects more urgent and vital to the war effort — radar and the atom bomb — were to occupy his mind for the next six years.

To make a radar that would be of practical use in the war — for example, as a means of detecting approaching enemy aircraft — it was necessary to increase greatly the power in the radio beam and to make it work at much shorter wavelengths. Oliphant adapted the klystron (used as a radiofrequency source in the cyclotron) for this purpose. But John Randall and Harry Boot, working in



Enthusiastic builder of particle accelerators and research centres

Oliphant's department, soon developed a much more efficient device, the resonant cavity magnetron. Within a few months it became the tool that helped to win the Battle of Britain in 1940 and avert Hitler's invasion of England. It was largely Oliphant's indefatigability that made this possible.

Oliphant played a somewhat similar role in the development of the atom bomb. Research on the bomb began in England in 1939, but the main impetus came from the calculations by Otto Frisch and Rudolf Peierls, early in 1940, which showed that the critical mass for a divergent chain reaction, propagated by fast neutrons in uranium-235, was only a few kilograms. It was Oliphant who brought the Frisch–Peierls memorandum to the attention of government authorities. As a result, the MAUD Committee was set up, charged with the development of the atom bomb.

Most of the experimental research on the physics of the bomb was done in Liverpool, where Frisch joined James Chadwick's team. By 1941, the scientific feasibility of the bomb had been established, but the separation of the uranium-235 isotope was too difficult a task for Britain under wartime conditions.

While on a trip to the United States on business connected with radar, in the autumn of 1941, Oliphant discovered that no work on the atom bomb was going on at

here, despite the fact that the report of the MAUD Committee had been sent to the relevant authorities. He immediately informed a few influential friends about the British findings. The outcome of Oliphant's indiscretion was the setting up of the Manhattan Project. He himself worked on the project, on the manufacture of uranium-235, in California and Tennessee.

Oliphant returned to Birmingham for only a few years before taking up, in 1950, permanent residence in his native country. During his time in Birmingham he designed the proton-synchrotron, in which particles could be accelerated to gigaelectronvolts (GeV). This work was based on the 'phase stability' principle, which he conceived before Vladimir Veksler and Edwin McMillan.

Oliphant was lured to Canberra by a promise from the Australian government to provide him with sufficient finances to set up a Research School of Physical Sciences at the new Australian National University. With his usual enthusiasm he threw himself into the project of creating a school of nuclear research equal to any in the world; this included a plan for an accelerator of novel design that would produce particles with a higher energy than achieved anywhere else. Although much of his planning came to fruition, he failed to build his ambitious 10-GeV 'cyclosynchrotron'; it was simply beyond the capacity of a small country, and his dream machine remained a "white Oliphant" (as it was described at the time). However, this failure did not diminish his high standing in Australia — he had the unique distinction, for a scientist, of serving as a governor of South Australia, the state of his birth.

Oliphant was vehemently opposed to the use of the atom bomb on the Japanese cities. He never overcame his feelings of guilt about his part in the Manhattan Project, and he frequently expressed his views to the Australian media. He also took an active part in international campaigns against nuclear weapons, particularly in the Pugwash Conferences on Science and World Affairs. One of the 22 participants in the first conference in 1957, he fully shared the Pugwash precept that scientists have a moral duty to be concerned about the social impact of their work. Describing himself as a 'belligerent pacifist', he advocated his conviction that war itself is evil and must not be tolerated by humanity.

Joseph Rotblat

Joseph Rotblat is at the Pugwash Conferences on Science and World Affairs, 63A Great Russell Street, London WC1B 3BJ, UK.
e-mail: pugwash@qmw.ac.uk

Maternal age and traits in offspring

The timing of a mouse's first litter influences the development of her pups.

We have investigated the effect of the age at first pregnancy in mice on maternal steroid hormone levels and how these influence the growth and sexual maturation of their pups. We find that early-adolescent and middle-aged pregnant mice have less serum oestradiol and a different pattern of serum testosterone compared with young-adult pregnant mice. Grown offspring of early-adolescent and middle-aged mothers have lower body weight and delayed puberty, and males have smaller reproductive organs than those produced by young adults. Female offspring themselves produce pups whose birthweight depends on the age at pregnancy of their grandmothers. Changes in a female's physiology during ageing can thus alter the growth and reproductive traits not only of her own offspring but also of subsequent generations.

We measured the concentration of steroid hormones in serum from pregnant CF-1 females between gestation day 16 and the last day of pregnancy, when the reproductive organs of fetuses are undergoing differentiation (Fig. 1). Serum testosterone was significantly lower in middle-aged than in either early-adolescent or young-adult pregnant females, and peaked significantly later (at gestation day 18) in the young-adult females. Serum oestradiol was significantly lower in early-adolescent and middle-aged pregnant females relative to the young adults, whereas serum progesterone did not differ significantly as a function of maternal age (Fig. 1a).

Offspring (F_1 generation) are designated by maternal age at mating in early adolescence (EA), young adulthood (YA) or middle age (MA). As young adults, YA males had significantly greater body weight and heavier testes and epididymides than EA or MA males (Fig. 1b). Maternal age seemed to influence testicular sperm production and storage, irrespective of its effect on the body weight of male offspring. The small size of MA male offspring was surprising, given that they were born in litters containing 40% fewer pups than YA males (Fig. 1a, legend) and all animals were delivered by caesarean section and reared by young-adult foster mothers.

Body weight at weaning (23 days old) of EA female offspring (12.1 ± 0.2 g) was significantly lower than for YA females (13.3 ± 0.3 g) and MA females (12.9 ± 0.4 g), which did not differ significantly. After being weaned, F_1 females were paired with a stud male until visibly pregnant. Based on the age at delivery of young and a 19-day pregnancy, the age at which the EA females

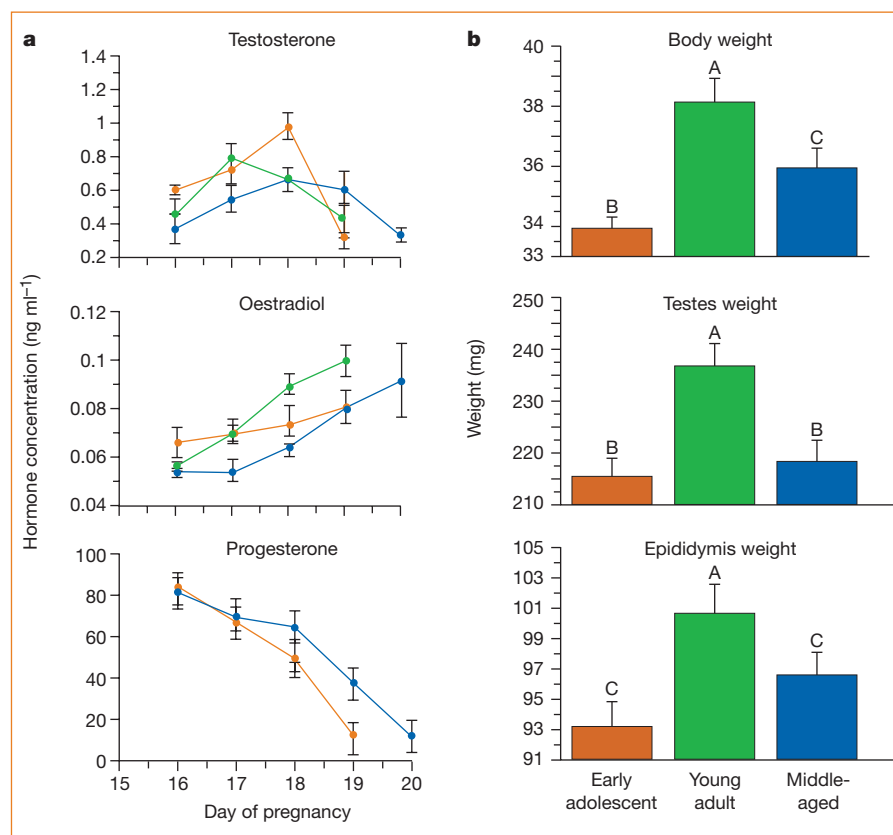


Figure 1 Maternal serum steroid-hormone concentrations and traits of male offspring. **a**, Pregnant mice were mated at 35 days old (early adolescent; red), at 3 months old (young adult; green), or at 9 months old (middle aged; blue). Maternal serum was collected on gestation day 16 (mating, day 0) through to the day of normal parturition. For middle-aged females, parturition was delayed by one day and this was associated with a delay (not statistically significant) in the decline in progesterone at the end of pregnancy. The number of live pups at birth differed significantly, with young-adult mothers producing more pups (12.1 ± 0.5) than early-adolescent females (10.3 ± 0.5), who produced more pups than middle-aged females (7.5 ± 0.5); data are means $\pm$ s.e.m.; n was 7–11 per group per day. **b**, Mean body weight ($\pm$ s.e.m.), and mean weight of testes and epididymides for 3-month-old male offspring (F_1 generation) of mothers of different ages. Early-adolescent, $n=25$; young-adult, $n=32$; and middle-aged, $n=23$. Body weight of males differs as a function of maternal age; the means for weight of testes and epididymides are adjusted for differences in body weight by analysis of covariance. A versus B, $P < 0.01$; A or B versus C, $P < 0.05$.

(35.2 ± 1.2 days old) and MA females (34.0 ± 1.3 days old) completed puberty (ovulated and mated) was significantly delayed relative to YA females (31.1 ± 0.8 days old). A small increase in oestradiol or other oestrogens during fetal life in female mice is also associated with advancement of puberty^{1,2}. Body weight and the age at puberty can both influence reproductive fitness and affect population dynamics³.

To investigate whether the effects of age at first pregnancy could be passed on to subsequent generations, we measured the body weight of the male and female pups (F_2 generation) produced by the F_1 females; litter size and sex ratio did not differ significantly. F_2 pups with young-adult grandmothers weighed 1.60 ± 0.01 g at birth and were significantly heavier than pups with grandmothers that had become pregnant in

either early adolescence (1.51 ± 0.01 g) or middle age (1.50 ± 0.01 g).

Maternal-age effects on human offspring have not been investigated, apart from those related to genetic abnormalities associated with ageing oocytes^{4,5}. Maternal differences in circulating oestradiol and testosterone associated with the timing of first pregnancy are also seen in rats⁶. The levels of oestradiol and testosterone in the mother during sexual differentiation in the fetus may be related to certain traits in human offspring, such as undescended testes and, subsequently, testicular cancer^{7–9}; these hormones *in utero* may permanently 'imprint' the function of cells in reproductive organs, the brain and many other tissues^{1,10,11}. The influence on offspring of age-related changes in the maternal reproductive and endocrine systems calls for further investigation as very

early (soon after puberty) and very late (approaching menopause) pregnancies are common in humans.

Ming-Hseng Wang, Frederick S. vom Saal

Division of Biological Sciences, University of Missouri, Columbia, Missouri 65211, USA
e-mail: vomsaalf@missouri.edu

1. vom Saal, F. S. *J. Anim. Sci.* **67**, 1824–1840 (1989).
2. Howdeshell, K. L., Hotchkiss, A. K., Thayer, K. A., Vandenberg, J. G. & vom Saal, F. S. *Nature* **401**, 763–764 (1999).
3. Clutton-Brock, T. H. (ed.) *Reproductive Success: Studies of Individual Variation in Contrasting Breeding Systems* (Univ.

- Chicago Press, Chicago, 1988).
4. vom Saal, F. S., Finch, C. E. & Nelson, J. F. in *Physiology of Reproduction* Vol. 2 (eds Knobil, E., Neill, J. & Pfaff, D.) 1213–1314 (Raven, New York, 1994).
5. Finch, C. E. & Kirkwood, T. B. L. *Chance, Development, and Aging* (Oxford Univ. Press, 2000).
6. Matt, D. W., Lee, J., Sarver, P. L., Judd, H. L. & Lu, J. K. H. *Biol. Reprod.* **34**, 478–487 (1986).
7. Bernstein, L. *et al. J. Natl Cancer Inst.* **76**, 1035–1039 (1986).
8. Bernstein, L. *et al. Br. J. Cancer* **58**, 379–381 (1988).
9. Key, T. *et al. Br. J. Cancer* **73**, 698–701 (1996).
10. Holliday, R. *Science* **238**, 163–170 (1987).
11. vom Saal, F. S. *et al. Proc. Natl Acad. Sci. USA* **94**, 2056–2061 (1997).

Intelligence

Maze-solving by an amoeboid organism

The plasmodium of the slime mould *Physarum polycephalum* is a large amoeba-like cell consisting of a dendritic network of tube-like structures (pseudopodia). It changes its shape as it crawls over a plain agar gel and, if food is placed at two different points, it will put out pseudopodia that connect the two food sources. Here we show that this simple organism has the ability to find the mini-

mum-length solution between two points in a labyrinth.

We took a growing tip of an appropriate size from a large plasmodium in a 25 × 35 cm culture trough and divided it into small pieces. We then positioned these in a maze created by cutting a plastic film and placing it on an agar surface. The plasmodial pieces spread and coalesced to form a single organism that filled the maze (Fig. 1a), avoiding the dry surface of the plastic film. At the start and end points of the maze, we placed 0.5 × 1 × 2 cm agar blocks containing nutrient (0.1 mg g⁻¹ of ground oat flakes). There were four possible routes (α 1,

α 2, β 1, β 2) between the start and end points (Fig. 1a).

The plasmodium pseudopodia reaching dead ends in the labyrinth shrank (Fig. 1b), resulting in the formation of a single thick pseudopodium spanning the minimum length between the nutrient-containing agar blocks (Fig. 1c). The exact position and length of the pseudopodium was different in each experiment, but the path through α 2 — which was about 22% shorter than that through α 1 — was always selected (Fig. 1d). About the same number of tubes formed through β 1 and β 2 as the difference (about 2%) in their path lengths is lost in the meandering of the tube trajectory and is within experimental error.

The addition of food leads to a local increase in the plasmodium's contraction frequency, initiating waves propagating towards regions of lower frequency^{1–5}, in accordance with the theory of phase dynamics⁶. The plasmodial tube is reinforced or decays when it lies parallel or perpendicular, respectively, to the direction of local periodic contraction⁷; the final tube, following the wave propagation, will therefore link food sites by the shortest path.

To maximize its foraging efficiency, and therefore its chances of survival, the plasmodium changes its shape in the maze to form one thick tube covering the shortest distance between the food sources. This remarkable process of cellular computation implies that cellular materials can show a primitive intelligence^{8–10}.

Toshiyuki Nakagaki*†, Hiroyasu Yamada*†‡, Ágota Tóth§

**Bio-Mimetic Control Research Center, RIKEN,*

Shimoshidami, Morioka, Nagoya 463-0003, Japan

e-mail: nakagaki@postman.riken.go.jp

†Local Spatio-Temporal Functions Laboratory, RIKEN, Wako 351-0198, Japan

‡Research Institute for Electronic Science,

Hokkaido University, Sapporo 060-0812, Japan

§Department of Physical Chemistry, University of Szeged, PO Box 105, Szeged H-6701, Hungary

1. Durham, A. C. & Ridgeway, E. B. *J. Cell Biol.* **69**, 218–223 (1976).
2. Matsumoto, K., Ueda, T. & Kobatake, Y. *J. Theor. Biol.* **122**, 339–345 (1986).
3. Miyake, Y., Tada, H., Yano, M. & Shimizu, H. *Cell Struct. Funct.* **19**, 363–370 (1994).
4. Nakagaki, T., Yamada, H. & Ito, M. *J. Theor. Biol.* **197**, 497–506 (1999).
5. Yamada, H., Nakagaki, T. & Ito, M. *Phys. Rev. E* **59**, 1009–1014 (1999).
6. Kuramoto, Y. in *Chemical Oscillations, Waves and Turbulence* (Springer, Berlin, 1984).
7. Nakagaki, T., Yamada, H. & Ueda, T. *Biophys. Chem.* **84**, 195–204 (2000).
8. Sepulchre, J. A., Babloyantz, A. & Steels, L. in *Proc. Int. Conf. on Artificial Neural Networks* (eds Kohonen, T. *et al.*) 1265–1268 (Elsevier, Amsterdam, 1991).
9. Sepulchre, J. A. & Babloyantz, A. *Phys. Rev. E* **48**, 187–195 (1993).
10. Steinbock, O., Tóth, Á. & Showalter, K. *Science* **267**, 868–871 (1995).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>).

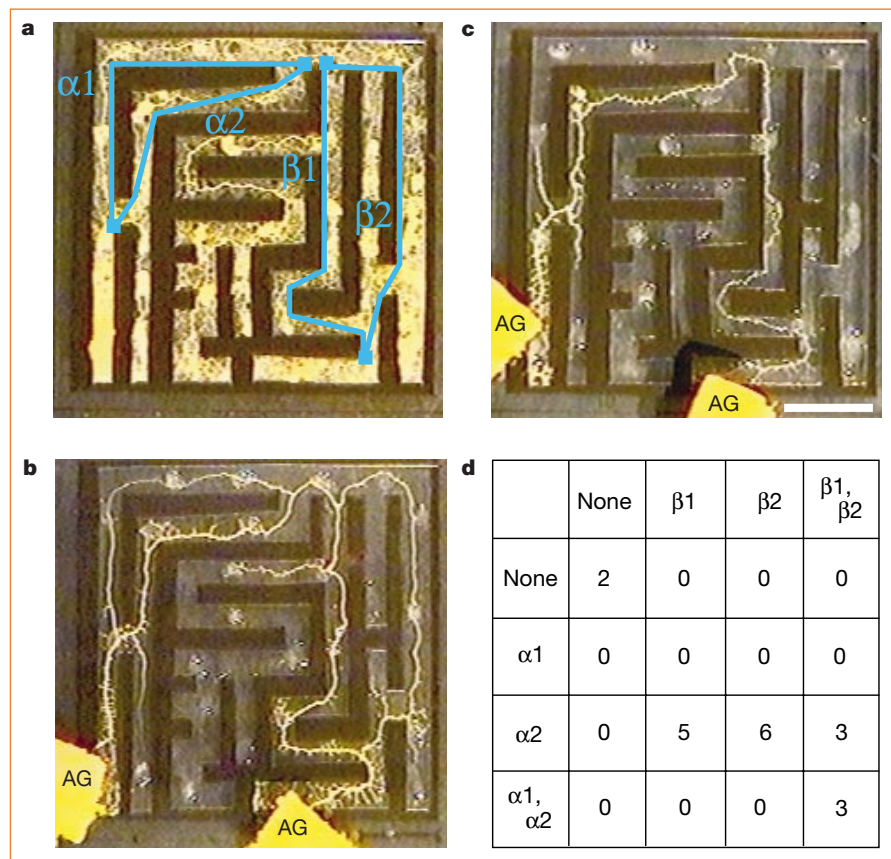


Figure 1 Maze-solving by *Physarum polycephalum*. **a**, Structure of the organism before finding the shortest path. Blue lines indicate the shortest paths between two agar blocks containing nutrients: α 1 (41 ± 1 mm); α 2 (33 ± 1 mm); β 1 (44 ± 1 mm); and β 2 (45 ± 1 mm). **b**, Four hours after the setting of the agar blocks (AG), the dead ends of the plasmodium shrink and the pseudopodia explore all possible connections. **c**, Four hours later, the shortest path has been selected. Plasmodium wet weight, 90 ± 10 mg. Yellow, plasmodium; black, 'walls' of the maze; scale bar, 1 cm. **d**, Path selection. Numbers indicate the frequency with which each pathway was selected. 'None', no pseudopodia (tubes) were put out. See Supplementary Information for an animated version of **a–c**.

RNA polymerase II elongation through chromatin

George Orphanides* & Danny Reinberg†

* Zeneca Central Toxicology Laboratory, Alderley Park, Cheshire, SK10 4TJ, UK

† Howard Hughes Medical Institute, Division of Nucleic Acids Enzymology, Department of Biochemistry, University of Medicine and Dentistry of New Jersey, Robert Wood Johnson Medical School, Piscataway, New Jersey 08854, USA

The machinery that transcribes protein-coding genes in eukaryotic cells must contend with repressive chromatin structures in order to find its target DNA sequences. Diverse arrays of proteins modify the structure of chromatin at gene promoters to help transcriptional regulatory proteins access their DNA recognition sites. The way in which disruption of chromatin structure at a promoter is transmitted through a whole gene has not been defined. Recent breakthroughs suggest that the passage of an RNA polymerase through a gene is coupled to mechanisms that propagate the breakdown of chromatin.

The need to compact 2 m of DNA into the nucleus of a eukaryotic cell has severe consequences for processes that require access to DNA. This compaction is achieved by the binding of proteins that mediate successive orders of DNA folding: two copies each of histones H2A, H2B, H3 and H4 form a protein core that wraps 146 base pairs of DNA tightly on its surface to form a nucleosome¹. With the aid of additional proteins, nucleosomes are compacted to form chromatin—a complex, highly ordered nucleoprotein assembly that is inaccessible to DNA-binding proteins². Decompaction of chromatin to facilitate access to DNA has been most widely studied for RNA polymerase II (RNAP II)-mediated transcription of protein-coding genes, a process that requires rapid access to genes for the response to environmental signals and programmed cellular events, but the underlying principles are equally applicable to any process requiring interaction with DNA.

Transcription can be divided into two basic stages: initiation, which involves the binding of transcription factors and RNAP II to specific sites in DNA (promoter regions) and the onset of RNA synthesis; and elongation, during which the polymerase tracks along the DNA and makes an RNA copy³. Until recently, interest in chromatin accessibility has focused on transcription initiation; however, initiation is thought only to result in the decompaction of chromatin surrounding the promoter, leaving transcribed regions in a compacted and repressive conformation. Recent reports suggest that eukaryotic cells are equipped with specialized proteins that help RNAP II pass through chromatin during transcription elongation. Here we discuss how these new findings implicate the elongation process in propagating the disruption of chromatin structure from the promoter region of a gene to the entire transcribed domain.

Proteins that decompact chromatin structure

Chromatin decompaction has been the subject of three decades of intense study. The lack of knowledge regarding the complex architecture of chromatin has hampered research in this area; however, it is clear that regions of the genome that are actively transcribed have a more open and accessible chromatin structure than non-transcribed regions⁴. Transcriptionally active, accessible regions are associated with a loss of proteins involved in the maintenance of higher-order chromatin structure. For example, histone H1, which binds to nucleosomes and promotes chromatin folding, is depleted in transcribed chromatin². Closer examination of the chromatin of transcribed regions reveals alterations in histone proteins themselves. Acetylation of lysine residues in the amino-terminal 'tails' of histones has long been correlated with transcriptional competence⁵. A good example of this phenomenon is found at the transcriptionally active β -globin locus of chicken erythrocytes, which contains 33 kilobases (kb) of accessible chromatin and is enriched in acetylated histones⁶. Structurally, histone tail acetylation disrupts histone—DNA and

inter-nucleosomal interactions by neutralizing positively charged lysine residues. The enzymes that catalyse these covalent modifications are the histone acetyltransferases (HATs) and histone deacetylases (HDACs), which add or remove an acetyl group, respectively⁷. The recent finding that many proteins that regulate transcription are, or can recruit, HATs and HDACs has reinforced the link between histone acetylation/deacetylation and gene activity.

Another class of factors that manipulate chromatin structure and have roles in transcriptional regulation are the chromatin-remodelling complexes, which use energy from ATP hydrolysis to disrupt chromatin and make DNA more accessible to DNA-binding proteins⁸. Chromatin remodelling enzymes have been purified from a variety of organisms, and most cells contain more than one type of complex^{1,8}. These complexes contain structurally related catalytic subunits, but differ in the way in which they manipulate chromatin. Most of these enzymes can alter the conformation of a nucleosome to increase access to DNA, and one complex, RSC⁹, can transfer an entire octamer of histone proteins from one region of DNA to another. Chromatin remodelling and HAT activities probably cooperate to overcome the transcriptional repression imposed by chromatin packaging.

RNA polymerase meets the nucleosome

RNAP II transcription is a complex process that is subject to chromatin repression at several stages (see Box 1). DNA binding by activator proteins is prevented by chromatin packaging¹⁰, but disruption of histone—DNA contacts by acetylation of histone tails can overcome this repression⁸. Similarly, the transient disruption of histone—DNA contacts by ATP-dependent chromatin-remodelling enzymes also facilitates DNA binding. Therefore, disruption of histone—DNA contacts by histone acetylation or chromatin remodelling allows DNA-binding proteins to compete with histones for DNA. Does this simple paradigm also apply to transcription elongation? Will weakening of histone—DNA contacts in a nucleosome allow its passage by an RNA polymerase? Initial experiments examining how an RNA polymerase copes with a nucleosome used small, single-subunit prokaryotic RNA polymerases from bacteriophages T7 and SP6 (refs 11, 12). Surprisingly, these small polymerases could transcribe completely through a nucleosome, albeit at a reduced rate compared with free DNA. Using a defined system containing a single nucleosome on a short linear DNA fragment, Felsenfeld and co-workers^{13,14} found that during elongation the entire octamer of histones is transferred backwards on the DNA fragment through a transiently formed DNA loop (Fig. 1). The observation that the histone octamer does not leave the DNA during the elongation process is consistent with the finding that histones remain associated with the DNA of genes being transcribed in the cell¹⁵. Although these studies have provided useful mechanistic insights, the templates used contained only a single nucleosome that is not subject to repressive inter-



nucleosome interactions that occur in natural chromatin. More physiological templates, containing long arrays of 15 or more nucleosomes, slow the progress of prokaryotic T7 RNA polymerase substantially, suggesting that inter-nucleosome contacts significantly repress elongation (for example, see ref. 16).

These experiments demonstrate that a single-subunit prokaryotic RNA polymerase can bypass a nucleosomal roadblock under certain conditions. Is this likely to be true for the large and cumbersome eukaryotic RNA polymerases with 12 or more protein subunits? The eukaryotic RNA polymerase III passes through a single nucleosome with pronounced pausing, suggesting that it is hindered by its large bulk¹⁷. Moreover, RNAP II is brought to an almost complete halt by arrays of nucleosomes, with strong polymerase-pausing sites in the nucleosome reflecting sites that induce natural pausing on free DNA^{18,19}. Furthermore, elongation factors that accelerate elongation on free DNA cannot overcome this chromatin block^{18,19}. The inability of eukaryotic RNA polymerases to elongate through nucleosomes *in vitro* may seem surprising, as they have evolved with their DNA template packaged into chromatin and, therefore, would be expected to have developed mechanisms to deal with nucleosomes. However, recent evidence indicates that in a cell these RNA polymerases recruit the assistance of cellular factors that can disrupt chromatin structure.

Helping RNAP II to elongate through chromatin

RNAP II in a cell travels at a rate of 25 nucleotides per second, an elongation rate that can only be achieved *in vitro* on free DNA templates (for example, see ref. 20). How does the polymerase achieve these rates in the repressive context of chromatin? This question has prompted researchers to seek conditions that will facilitate transcription elongation through chromatin *in vitro*. Chromatin remodelling by the SWI/SNF complex can promote RNAP II elongation through a nucleosome²¹, presumably by disrupting histone–DNA interactions. Similarly, the chromatin-associated HMG14 protein, which is found in the chromatin of active genes *in vivo*²², can modestly enhance RNAP II elongation through chromatin *in vitro*²³.

Box 1

Transcription of protein-coding genes is a complex process

In its most general form, transcription activation by RNA polymerase II (RNAP II) begins with the binding of activator proteins to DNA sites adjacent to the start site of transcription (the promoter region). Once bound, the activators can recruit RNAP II in a step involving a set of accessory proteins—the general transcription factors (GTFs)—that accurately position the polymerase over the transcription start site^{3,49}. RNAP II then begins transcription, clears the promoter and enters the elongation phase. Finally, directed by DNA sequences at the end of a gene, RNAP II terminates transcription and releases newly synthesized RNA. Transcription is regulated primarily at initiation, but a set of factors can stimulate the elongation process by suppressing transient polymerase pausing or by increasing the rate of RNA polymerization⁵⁰.

Soon after initiating transcription, RNAP II is hyperphosphorylated on its C-terminal domain (CTD), an unusual structure composed of several repeats of a heptapeptide motif. Mounting evidence suggests that CTD phosphorylation regulates the interaction of RNAP II with proteins involved in transcriptional regulation and mRNA maturation⁵¹. RNAP II containing an unphosphorylated CTD is bound by the mediator complex, a multi-protein assembly involved in the recruitment of RNAP II to active promoters⁵². Once elongation is underway, the polymerase breaks its interactions with transcription initiation factors and its CTD becomes a docking site for enzymes that catalyse RNA polyadenylation, capping and splicing.

Another complex that can facilitate RNAP II elongation through nucleosomes is FACT. When added to a completely defined transcription system using chromatin as the template, the heterodimeric FACT complex enables RNAP II to elongate through nucleosomes¹⁹. One of the subunits of FACT is a human homologue of the *Saccharomyces cerevisiae* Spt16 protein, which has been implicated in modulating chromatin structure for transcription elongation by genetic experiments^{24–26}. The smaller subunit of FACT is the HMG-1-like protein SSRP1. Biochemical analysis suggests that FACT stably and specifically interacts with histones H2A and H2B, and covalent crosslinking of histones in a nucleosome, to prevent removal of these histones, abrogates FACT activity²⁶. These properties suggest that FACT disrupts nucleosomes during RNAP II elongation by transiently binding and removing histones H2A and H2B. Consistent with this model, yeast strains with mutations in histone H4, which alter the interaction of histones H2A and H2B with other histones in the nucleosome, exhibit identical phenotypes to strains carrying mutations in the Spt16 subunit of FACT²⁷. Furthermore, chromatin that contains transcribed sequences is deficient in histones H2A and H2B and is preferentially bound by RNAP II (ref. 28).

Another group of proteins implicated in relieving the chromatin block to transcription are the Spt4, Spt5 and Spt6 proteins. Yeast strains with mutations in the genes encoding these proteins exhibit phenotypes consistent with defects in transcription elongation²⁹ and share many phenotypes with strains containing mutations in the Spt16 subunit of FACT and in histone proteins³⁰. A human complex of Spt4 and Spt5 proteins binds to RNAP II and modulates its elongation activity on naked DNA templates *in vitro*³¹. In addition, this complex, called DSIF, can promote RNAP II elongation through chromatin templates assembled *in vitro* (T. Wada, G.O., H. Handa and D.R., unpublished data). The third member of this group of proteins, Spt6, can bind to histones and alters chromatin structure *in vitro*³². The roles played by FACT, the SWI/SNF complex, HMG14 and Spt4, Spt5 and Spt6 in facilitating elongation through chromatin

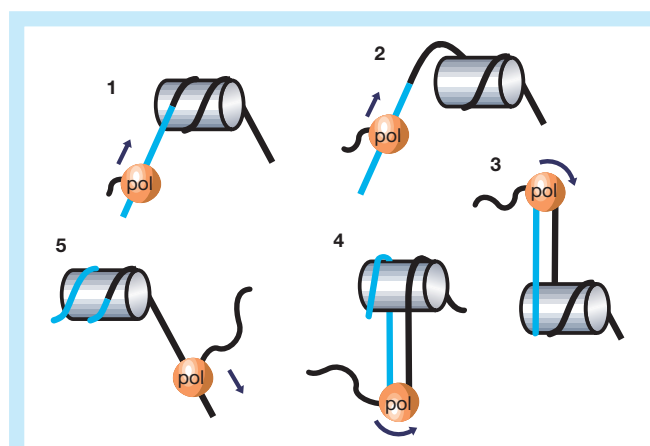


Figure 1 Model for transcription through a nucleosome by a single-subunit prokaryotic RNA polymerase^{13,14}. The RNA polymerase approaches the nucleosome and begins to synthesize a transcript (1; the DNA present in the nucleosome is shaded black). Transcription of the first 25 bp of DNA in the nucleosome is rapid and results in the displacement of DNA from the central histone octamer (2). The DNA segment behind the polymerase binds to the freshly exposed surface of the histone octamer, forming a DNA loop (3). Further polymerase elongation is hindered by this tight loop, which prevents the polymerase from rotating around the DNA as it reads the nucleotide sequence. The DNA behind the polymerase then transiently dissociates from the histone octamer, breaking the loop and allowing further elongation. This results in progression of the loop further into the nucleosome (4). This cycle of loop formation and breaking is repeated until the polymerase passes completely through the nucleosome (5; for details, see ref. 14). This process results in the transfer of the entire octamer of histone proteins backward on the same DNA segment. Figure adapted with permission from ref. 48.

in a cell are unknown, but are likely to be determined through a combination of genetic and biochemical approaches.

Hitching a ride on RNAP II

If the factors described in the previous section facilitate elongation through chromatin in a cell, how are they likely to be targeted to transcribed regions of the genome? It is established that activities that modify chromatin structure can be recruited to promoter regions to facilitate transcription initiation through direct interactions with DNA-bound activator proteins⁸, but how are they directed to downstream regions to facilitate elongation? One way in which these activities could gain access to the entire transcribed region is by hitching a ride on the polymerase as it travels on its journey through a gene³³. In this way, the chromatin modifier would gain access to, and could promote the disruption of, the whole transcribed region (Fig. 2).

For this targeting mechanism to be effective, the modifying activities must be able to recognize and bind to polymerases that are elongating, and not to polymerases that are free in the nucleus or are at

gene promoters. It is likely that the 'tag' that distinguishes an elongating polymerase is phosphorylation of its carboxy-terminal domain (CTD) tail (see Box 1). The first evidence that a chromatin-remodelling activity can recognize and bind to an elongating polymerase was the finding that the PCAF HAT, which acetylates histones H3 and H4 in a nucleosome³⁴, binds specifically to the phosphorylated, elongating form of RNAP II (ref. 35). More recently, the Svejstrup laboratory^{36,37} has characterized the composition of the elongating *S. cerevisiae* RNAP II complex. This resulted in the isolation of a heterotrimeric complex, called 'elongator', that associates only with the phosphorylated, elongating form of RNAP II and can also be found in a free form^{36,37}. The 60-kDa subunit of the elongator is the Ebp3 protein, which contains HAT activity towards all four histones. These characteristics of elongator suggest that it is recruited to elongating RNAP II to acetylate the histones of transcribed chromatin. Genetic experiments reveal that the elongator complex is dispensable for yeast survival, but has an important function in the induction of certain genes.

Transcription elongation and histone acetylation

Although the neutralization of negatively charged histone tails by acetylation of lysines can disrupt chromatin structure, there is evidence that this is not how the transcriptional effects of histone acetylation are propagated in a cell. For example, the observation that acetylation of histone H3 is associated with both transcriptional activation and the generation of repressed chromatin structures suggests that charge neutralization by acetylation is not the direct cause of transcriptional activation³⁸. A more appealing model for the function of histone acetylation proposes that the pattern of histone tail modifications serves as a recognition code for the binding of proteins that modulate chromatin structure (the 'histone code' hypothesis³⁸). Therefore, acetylation of different histone tail lysines by one or more HATs travelling with elongating RNAP II may provide a combination of recognition sites for chromatin-modifying complexes that assist the elongation process, such as FACT, HMG14 and additional HAT complexes.

The overall picture that emerges is one in which the modification of chromatin by HATs that track with RNAP II leads to derepression of a whole transcription unit. This model also implies a mechanism for the way in which active transcription can be turned off. Competition with HDACs that are free in the nucleus means that maintaining histones in an acetylated state requires constant transcription, a proposal supported by experiments showing that the establishment of an unfolded chromatin domain *in vivo* requires transcription elongation and histone acetylation³⁹. The state of histone tail acetylation is a dynamic equilibrium determined by the activities of HATs bound to elongating RNAP II and HDACs. Once the RNAP II traffic along a gene is decreased, which is governed by signals at the promoter, the equilibrium shifts in favour of the HDACs. Loss of histone tail acetylation may then result in the rapid conversion of chromatin structure to a repressed conformation.

Is a 'pioneer' polymerase required?

Activities such as chromatin-remodelling complexes, HATs and FACT can disrupt chromatin structure at the level of the nucleosome to facilitate transcription elongation *in vitro*. But the nucleosome is only the first level in chromatin compaction. The precise degree of DNA compaction faced by elongating RNAP II in a cell is not known and, therefore, it is unclear whether these activities are sufficient for elongation through chromatin *in vivo*. Two models, which differ in the extent of chromatin decompaction that occurs following the binding of activators to promoters, are possible (Fig. 3). In the first model, recruitment of chromatin-modifying activities by activators results in complete decompaction of chromatin surrounding the activator-binding sites and only partial decompaction elsewhere in the gene. In this case, elongating RNAP II faces a compacted chromatin template. In the second model, activators promote decom-

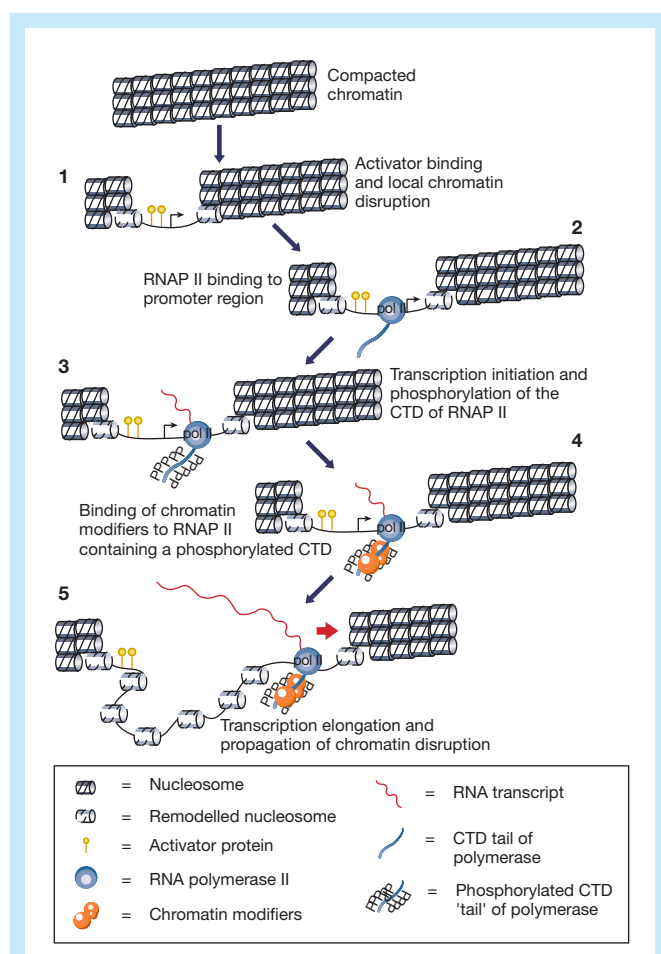


Figure 2 Propagation of chromatin disruption by chromatin-modifying activities that track with elongating RNAP II. One way in which chromatin-modifying activities can gain access to the whole transcribed region is by binding to RNAP II and travelling with it during elongation. Binding of activator proteins to the promoter results in recruitment of chromatin-modifying activities and local chromatin disruption (1). With the assistance of the general transcription factors, RNAP II then binds to the promoter region (2). Shortly after the initiation of transcription, the C-terminal domain (CTD) of RNAP II is hyperphosphorylated (3). RNAP II with a hyperphosphorylated CTD is recognized and bound by chromatin-modifying factors such as PCAF and/or elongator (4). The chromatin-modifying activities then travel with elongating RNAP II, leading to the propagation of chromatin disruption (5). This mechanism would enable the chromatin-modifying activities to gain access to the entire transcribed region and would result in the propagation of chromatin disruption through the entire length of the gene. For illustrative purposes only, the template ahead of the polymerase is shown here as a compacted chromatin fibre.

paction of chromatin over the whole gene. In this case, the elongating RNAP II would find partially decompacted nucleosomes in its path (Fig. 3).

Very little experimental evidence exists to support either of these models. If the first one is correct, RNAP II must penetrate and unpack a repressive chromatin fibre. To facilitate this, it is possible that the first polymerase to transcribe a gene is a specialized 'pioneer' polymerase equipped with additional tools that break down higher-order chromatin structures. Histone acetyltransferases that travel with subsequent elongating polymerases would then function to maintain transcribed chromatin in an accessible conformation. If a pioneer RNAP II complex exists, it is likely to be in low abundance and, therefore, may be difficult to purify by conventional means. Nevertheless, experiments in *S. cerevisiae* suggest that the first polymerase to transcribe a given gene is different from subsequent polymerases. Upon transcription initiation, the mediator complex of *S. cerevisiae* RNAP II, which is associated with a subset of RNAP II complexes in the cell, is left at the promoter, suggesting that subsequent polymerases loaded onto the gene at the promoter do not require associated mediator complexes (S. Hahn, personal communication). A prediction of this model is that the pioneer polymerase, faced with the repressive, compacted chromatin fibre, will elongate more slowly than will the polymerases that follow. This does not seem to be the case: on the *Drosophila* HSP70 gene, the first polymerase synthesizes a transcript at a rate equal to subsequent polymerases⁴⁰. Unlike most promoters, however, the unactivated form of the HSP70 promoter contains a transcriptionally engaged, paused RNAP II, and one cannot rule out that a pioneer polymerase has already transcribed the gene to establish an accessible chromatin conformation. Moreover, the transcription complex formed at heat shock promoters, such as the HSP70 gene promoter, appears to be different from that formed on most promoters^{41,42}. Thus, further definition of the conformation of the chromatin template faced by elongating RNAP II requires direct examination of the structure of transcribed chromatin subsequent to transcription initiation, both before and after elongation.

Does intergenic transcription disrupt chromatin

In some cases, chromatin modification and disruption are not restricted to promoters and genes, but are spread over complete gene loci. For example, the entire 33 kb of the chicken β -globin locus, including intergenic regions, contains disrupted chromatin with acetylated histones⁶. The human β -globin locus consists of five developmentally regulated genes whose expression is controlled by a stretch of DNA known as the locus control region (LCR). The LCR contains clusters of transcription-factor-binding sites and can promote transcription of genes in an adjacent chromatin region when placed at any chromosomal location⁴³. Two hypotheses have been put forward to explain the way in which an open chromatin conformation is established at the β -globin locus to facilitate transcription. The first proposes that proteins bound to the LCR recruit factors that disrupt chromatin over the entire β -globin locus; the second hypothesis directly implicates transcription elongation in generating an open chromatin conformation^{33,44}.

If transcription elongation is directly involved in disrupting chromatin at the β -globin locus, one would predict the existence of large intergenic transcripts in cells containing active β -globin loci. Strikingly, large and rare intergenic transcripts derived from the human β -globin locus have been detected⁴⁴. Recently, Frazer and co-workers⁴⁵ examined the link between intergenic transcription and chromatin disruption and found the locus to be divided into three distinct chromatin domains, each of which corresponds to a specific intergenic transcript. A tight temporal correlation exists between the synthesis of these transcripts and chromatin structural changes associated with activation of specific genes in the locus. Moreover, the intergenic transcripts appear to originate from precise regions, at least one of which has been implicated in chromatin disruption. These observations have challenged the long-standing hypothesis that the LCR of the β -

globin locus is necessary and sufficient for chromatin disruption. Indeed, recent studies have demonstrated that the LCR of the mouse β -globin locus is not required for chromatin disruption, but is required for high-level transcription of the β -globin genes⁴⁶. It will be interesting to determine whether intergenic transcription is involved in chromatin disruption at other gene loci and whether a distinct form of the polymerase, or a pioneer polymerase, is responsible.

Future directions

Once thought to be an inert structural feature, chromatin is now known to be an integral part of the machinery that controls gene expression. The progress described here sets the stage for the next decade of experiments aimed at understanding how chromatin is disrupted and re-structured during transcription. Future work will reveal the molecular details of these chromatin changes, including the pattern of histone acetylation, changes in nucleosome conformation and the binding of accessory proteins. Attempts must be made to determine the role of histone tail acetylation by HAT activities that track with RNAP II. This includes the mapping of histone tail lysine residues acetylated during elongation, and the identification and characterization of proteins that recognize these modifications. Engineering strains of the yeast *S. cerevisiae* with mutations that inactivate these HAT activities, or replace their target lysine

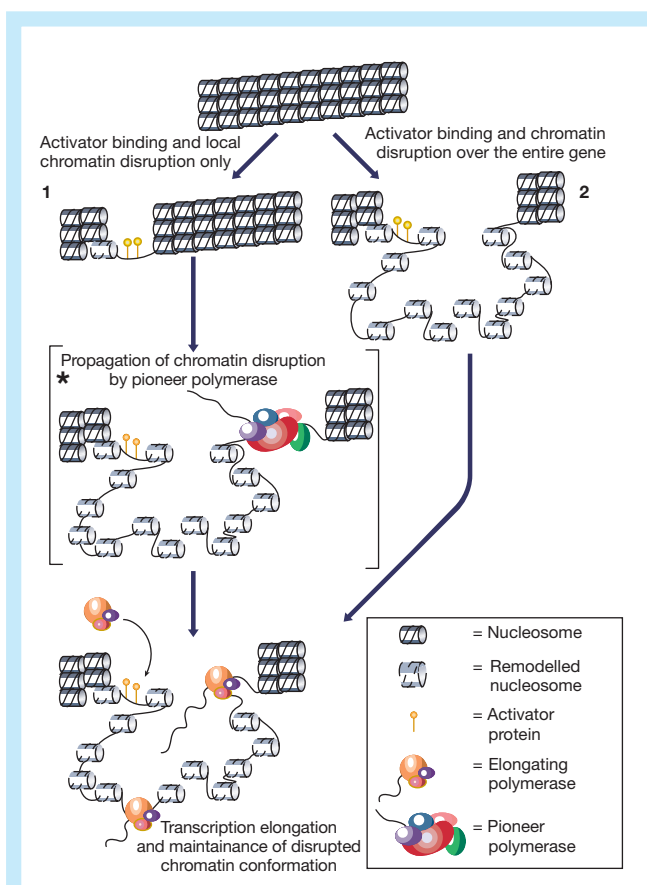


Figure 3 Models for the order of events in the decompaction of transcribed chromatin. In model 1, activator proteins bound to promoter regions promote only a local decompaction of chromatin surrounding their DNA-binding sites. This leaves the remainder of the gene in a compacted or partially decompacted conformation. In this case, the elongating RNAP II will face a repressive chromatin structure. If this model is correct, disruption of transcribed chromatin may be accomplished by a dedicated 'pioneer' polymerase equipped with tools to penetrate and decompact higher-order chromatin structure (asterisk). Transcribed chromatin may then be maintained in an accessible conformation by HAT complexes associated with subsequent polymerases. In an alternative model (2), activators promote chromatin decompaction over the entire gene. In this case, RNAP II will find only nucleosomes in its path.

residues in histones with inert amino acids, will be crucial in unravelling these problems. The histone proteins of chromatin are subject to a plethora of covalent modifications, including phosphorylation and methylation, and it remains to be seen whether other histone modifications are involved in transcription elongation. It is unlikely that a general mechanism exists for the derepression of chromatin for transcription: for some genes (for example, those regulated by the GCN5 HAT⁴⁷), histone acetylation appears to be restricted to the promoter region. Deciphering the underlying reasons for this diversity will be a major task. Biochemical studies have identified factors (for example, FACT, HMG14 and remodelling complexes) that help RNAP II elongate through chromatin. The observation that, even in the presence of these factors, elongation by RNAP II proceeds much more slowly *in vitro* than it does *in vivo*, suggests that other critical factors remain to be discovered.

Our progress in understanding mechanisms that regulate chromatin dynamics is limited by the lack of information regarding the overall organization of compacted chromatin; we cannot fully comprehend how chromatin is taken apart unless we first know how it is put together. Currently, our knowledge of chromatin decompaction extends only to the disruption of nucleosomes—the first level of chromatin compaction. The DNA template goes through many more degrees of folding before a mature chromatin fibre is formed, and we have yet to discover the machinery that mediates this folding. Transcription is proving to be an ideal model system for studying transitions in chromatin structure. It is likely that the lessons learned from transcription will be relevant to other processes that use DNA as a template and will have implications for diverse biological processes. □

- Kornberg, R. D. & Lorch, Y. Twenty-five years of the nucleosome, fundamental particle of the eukaryotic chromosome. *Cell* **98**, 285–294 (1999).
- van Holde, K. E. Chromatin (Springer, New York, 1989).
- Orphanides, G., Lagrange, T. & Reinberg, D. The general transcription factors of RNA polymerase II. *Genes Dev.* **10**, 2657–2683 (1996).
- Gross, D. S. & Garrard, W. T. Nuclease hypersensitive sites in chromatin. *Annu. Rev. Biochem.* **57**, 159–157 (1988).
- Struhl, K. Histone acetylation and transcriptional regulatory mechanisms. *Genes Dev.* **12**, 599–606 (1998).
- Hebbes, T. R., Clayton, A. L., Thorne, A. W. & Crane-Robinson, C. Core histone hyperacetylation co-maps with generalized DNase I sensitivity in the chicken beta-globin chromosomal domain. *EMBO J.* **13**, 1823–1830 (1994).
- Kuo, M.-H. & Allis, C. Roles of histone acetyltransferases and deacetylases in gene regulation. *BioEssays* **20**, 615–626 (1998).
- Kingston, R. E. & Narlikar, G. J. ATP-dependent remodelling and acetylation as regulators of chromatin fluidity. *Genes Dev.* **13**, 2339–2352 (1999).
- Lorch, Y., Zhang, M. & Kornberg, R. D. Histone octamer transfer by a chromatin-remodelling complex. *Cell* **96**, 389–392 (1999).
- Owen-Hughes, T. & Workman, J. L. Experimental analysis of chromatin function in transcription control. *Crit. Rev. Euk. Gene Exp.* **4**, 403–441 (1994).
- Lorch, Y., LaPointe, J. W. & Kornberg, R. D. Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones. *Cell* **49**, 203–210 (1987).
- Clark, D. J. & Felsenfeld, G. A nucleosome core is transferred out of the path of a transcribing polymerase. *Cell* **71**, 11–22 (1992).
- Studitsky, V. M., Clark, D. J. & Felsenfeld, G. A histone octamer can step around a transcribing polymerase without leaving the template. *Cell* **76**, 371–382 (1994).
- Studitsky, V. M., Clark, D. J. & Felsenfeld, G. Overcoming a nucleosomal barrier to transcription. *Cell* **83**, 19–27 (1995).
- Nacheva, G. A. *et al.* Change in the pattern of histone binding to DNA upon transcriptional activation. *Cell* **58**, 27–36 (1989).
- O'Neill, T. E., Roberge, M. & Bradbury, E. M. Nucleosome arrays inhibit both initiation and elongation of transcripts by bacteriophage T7 RNA polymerase. *J. Mol. Biol.* **223**, 67–78 (1992).
- Studitsky, V. M., Kassavetis, G. A., Geiduschek, E. P. & Felsenfeld, G. Mechanism of transcription through the nucleosome by eukaryotic RNA polymerase. *Science* **278**, 1960–1963 (1997).
- Izban, M. G. & Luse, D. S. Transcription on nucleosomal templates by RNA polymerase II in vitro: inhibition of elongation with enhancement of sequence-specific pausing. *Genes Dev.* **5**, 683–696 (1991).
- Orphanides, G., LeRoy, G., Chang, C.-H., Luse, D. S. & Reinberg, D. FACT, a factor that facilitates transcript elongation through nucleosomes. *Cell* **92**, 105–116 (1998).
- Izban, M. J. & Luse, D. S. Factor-stimulated RNA polymerase II transcribes at physiological elongation rates on naked DNA but very poorly on chromatin templates. *J. Biol. Chem.* **267**, 13647–13655 (1992).
- Brown, S. A., Imbalzano, A. N. & Kingston, R. E. Activator-dependent regulation of transcriptional pausing on nucleosomal templates. *Genes Dev.* **10**, 1479–1490 (1996).
- Bustin, M. & Reeves, R. High-mobility-group chromosomal proteins: architectural components that facilitate chromatin function. *Prog. Nucleic Acid Res. Mol. Biol.* **54**, 35–100 (1996).
- Ding, H. F., Rimsky, S., Batson, S. C., Bustin, M. & Hansen, U. Stimulation of RNA polymerase II elongation by chromosomal protein HMG-14. *Science* **265**, 796–799 (1994).
- Rowley, A., Singer, R. A. & Johnston, G. C. CDC68, a yeast gene that affects regulation of cell proliferation and transcription, encodes a protein with a highly acidic carboxyl terminus. *Mol. Cell. Biol.* **11**, 5718–5726 (1991).
- Malone, E. A., Clarke, C. D., Chiang, A. & Winston, F. Mutations in SPT16/CDC68 suppress *cis*- and *trans*-acting mutations that affect promoter function in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **11**, 5710–5717 (1991).
- Orphanides, G., Wu, W. H., Lane, W. S., Hampsey, M. & Reinberg, D. The chromatin-specific transcription elongation factor FACT comprises human SPT16 and SSRP1 proteins. *Nature* **400**, 284–288 (1999).
- Santisteban, M. S., Arents, G., Moudrianakis, E. N. & Smith, M. M. Histone octamer function in vivo: mutations in the dimer-tetramer interfaces disrupt both gene activation and repression. *EMBO J.* **16**, 2493–2506 (1997).
- Baer, B. W. & Rhodes, D. Eukaryotic RNA polymerase II binds to nucleosome cores from transcribed genes. *Nature* **301**, 482–488 (1983).
- Hartzog, G. A., Wada, T., Handa, H. & Winston, F. Evidence that Spt4, Spt5, and Spt6 control transcription elongation by RNA polymerase II in *Saccharomyces cerevisiae*. *Genes Dev.* **12**, 357–369 (1998).
- Winston, F. & Carlson, M. Yeast SNF/SWI transcriptional activators and the SPT/SIN chromatin connection. *Trends Genet.* **8**, 387–391 (1992).
- Wada, T. *et al.* DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. *Genes Dev.* **12**, 343–356 (1998).
- Bortvin, A. & Winston, F. Evidence that Spt6 controls chromatin structure by a direct interaction with histones. *Science* **272**, 1473–1476 (1996).
- Travers, A. Chromatin modification by DNA tracking. *Proc. Natl Acad. Sci. USA* **96**, 13634–13637 (1999).
- Yang, X. J., Ogryzko, V. V., Nishikawa, J., Howard, B. H. & Nakatani, Y. A p300/CBP-associated factor that competes with the adenoviral oncoprotein E1A. *Nature* **382**, 319–324 (1996).
- Cho, H. *et al.* A human RNA polymerase II complex containing factors that modify chromatin structure. *Mol. Cell. Biol.* **18**, 5355–5363 (1998).
- Otero, G. *et al.* Elongator, a multisubunit component of a novel RNA polymerase II holoenzyme for transcriptional elongation. *Mol. Cell* **3**, 109–118 (1999).
- Wittschleben, B. O. *et al.* A novel histone acetyltransferase is an integral subunit of elongating RNA polymerase II holoenzyme. *Mol. Cell* **4**, 123–128 (1999).
- Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* **403**, 41–45 (2000).
- Walia, H., Chen, H. Y., Sun, J. M., Holth, L. T. & Davie, J. R. Histone acetylation is required to maintain the unfolded nucleosome structure associated with transcribing DNA. *J. Biol. Chem.* **273**, 14516–14522 (1998).
- O'Brien, T. & Lis, J. T. Rapid changes in Drosophila transcription after an instantaneous heat shock. *Mol. Cell. Biol.* **13**, 3456–3463 (1993).
- Lee, D. & Lis, J. T. Transcriptional activation independent of TFIIF kinase and the RNA polymerase II mediator *in vivo*. *Nature* **393**, 389–392 (1998).
- McNeil, J. B., Agah, H. & Bentley, D. Activated transcription independent of the RNA polymerase II holoenzyme in budding yeast. *Genes Dev.* **12**, 2510–2521 (1998).
- Fraser, P. & Grosfeld, F. Locus control regions, chromatin activation and transcription. *Curr. Opin. Cell Biol.* **10**, 361–365 (1998).
- Ashe, H. L., Monks, J., Wijgerde, M., Fraser, P. & Proudfoot, N. J. Intergenic transcription and transduction of the human β -globin locus. *Genes Dev.* **11**, 2494–2509 (1997).
- Gribnau, J., Diderich, K., Pruzina, S., Calzolari, R. & Fraser, P. Intergenic transcription and developmental remodelling of chromatin subdomains in the human β -globin locus. *Mol. Cell* **5**, 377–386 (2000).
- Bender, M. A., Bulger, M., Close, J. & Groudine, M. β -globin switching and DNase I sensitivity of the endogenous β -globin locus in mice do not require the locus control region. *Mol. Cell* **5**, 387–393 (2000).
- Kuo, M.-H., Zhou, J., Jambeck, P., Churchill, M. E. A. & Allis, C. D. Histone acetyltransferase activity of yeast Gcn5p is required for activation of target genes *in vivo*. *Genes Dev.* **12**, 627–639 (1998).
- Felsenfeld, G., Boyes, J., Chung, J., Clark, D. & Studitsky, V. Chromatin structure and gene expression. *Proc. Natl Acad. Sci. USA* **93**, 9384–9388 (1996).
- Roeder, R. G. The role of the general initiation factors in transcription by RNA polymerase II. *Trends Biochem. Sci.* **21**, 327–335 (1996).
- Uptain, S. M., Kane, C. M. & Chamberlin, M. J. Basic mechanisms of transcript elongation and its regulation. *Annu. Rev. Biochem.* **66**, 117–172 (1997).
- Neugebauer, K. M. & Roth, M. B. Transcription units as RNA processing units. *Genes Dev.* **11**, 3279–3285 (1997).
- Koleske, A. J. & Young, R. A. The RNA polymerase II holoenzyme and its implications for gene regulation. *Trends Biochem. Sci.* **20**, 113–116 (1995).

Acknowledgements

We thank M. Hampsey, D. Luse and S. Hahn for helpful comments, and R. Sternglanz, J. Workman and S. Hahn for communication of unpublished results. Work in D.R.'s laboratory is supported by grants from the NIH and the Howard Hughes Medical Institute.

Correspondence should be addressed to D.R. (e-mail: reinbed@umdj.edu).



A yeast prion provides a mechanism for genetic variation and phenotypic diversity

Heather L. True & Susan L. Lindquist

Department of Molecular Genetics and Cell Biology, Howard Hughes Medical Institute, The University of Chicago, Chicago, Illinois 60637, USA

A major enigma in evolutionary biology is that new forms or functions often require the concerted effects of several independent genetic changes. It is unclear how such changes might accumulate when they are likely to be deleterious individually and be lost by selective pressure. The *Saccharomyces cerevisiae* prion $[PSI^+]$ is an epigenetic modifier of the fidelity of translation termination, but its impact on yeast biology has been unclear. Here we show that $[PSI^+]$ provides the means to uncover hidden genetic variation and produce new heritable phenotypes. Moreover, in each of the seven genetic backgrounds tested, the constellation of phenotypes produced was unique. We propose that the epigenetic and metastable nature of $[PSI^+]$ inheritance allows yeast cells to exploit pre-existing genetic variation to thrive in fluctuating environments. Further, the capacity of $[PSI^+]$ to convert previously neutral genetic variation to a non-neutral state may facilitate the evolution of new traits.

How might several independent genetic changes occur in concert to produce a new form or function? One solution is to posit the existence of mechanisms that allow changes to accumulate in a temporarily neutral fashion. Such a mechanism was proposed in 1972 when Koch¹ extended Ohno's recognition of the importance of gene duplications in the evolution of new functions² to include inactivation, modification and reactivation. A duplicated gene that is retained in an inactive state would be free to accumulate variations that might compromise fitness, and beneficial combinatorial change could occur upon reactivation¹. Gene duplications are common and these genes frequently acquire divergent functions^{3–5}, indicating that such a mechanism may have operated widely in evolution. However, the known mechanisms for the reactivation of inactive genes work sporadically, act infrequently and provide no obvious means for sampling coding changes in several genes simultaneously⁶.

Here we report that a protein-based element of inheritance (a prion) in the yeast *Saccharomyces cerevisiae* provides a means to expose this and other types of silent protein-coding information. The prion $[PSI^+]$ reduces the fidelity with which ribosomes terminate translation at stop codons in a metastable, heritable manner^{7–9}. The reduction in translational fidelity caused by $[PSI^+]$ is routinely monitored by the production of active products from genetic markers containing stop-codon mutations⁷. Presumably, $[PSI^+]$ also causes ribosomes to read through some naturally occurring stop-codons, but its biological significance is a mystery.

Despite this uncertainty, a great deal of information has accumulated on the mechanism of $[PSI^+]$ inheritance because it represents a new class of genetic elements that produce heritable changes in phenotype with no underlying changes in nucleic acids¹⁰. $[PSI^+]$ propagates through a self-perpetuating change in the state of Sup35, a translation termination factor^{11–14}. In its prion state, Sup35 forms inactive complexes that have the capacity to capture newly made Sup35 and convert it to the same form, perpetuating the loss-of-function phenotype¹³.

Sup35 has three distinct regions^{15–17}. The carboxyl-terminal region (C) is responsible for the translation-termination activity and is essential for viability¹⁸. The amino-terminal (N) and middle (M) regions allow Sup35 to acquire a stable prion conformation and switch between $[PSI^+]$ and $[psi^-]$ states^{13,14,18,19}. The N and M regions can be deleted in cells that do not carry the prion, $[psi^-]$ cells, with no apparent effect¹⁸. Deletion of N and M in $[PSI^+]$ cells results in

the loss of the prion and restoration of translational fidelity¹⁸. The prion-determining region of Sup35 has been retained in distantly related yeast^{20–23}, indicating that the unusual ability of Sup35 to produce a heritable conformational and phenotypic switch may provide a selective advantage.

Diverse growth phenotypes produced by $[PSI^+]$

We compared the growth characteristics of $[PSI^+]$ and $[psi^-]$ cells in more than 150 phenotypic assays (Table 1, Supplementary Information, and data not shown) including growth on fermentable and non-fermentable carbon sources (Figs 1, 2a and 3), on simple and complex nitrogen sources in the presence of salts and metals (Fig. 2b), in the presence of inhibitors of diverse cellular processes, such as DNA replication, signalling, protein glycosylation, and microtubule dynamics (Fig. 2c), under general stress conditions, and at different temperatures (Fig. 4). We examined cells from seven different genetic backgrounds to distinguish traits that might be a universal feature of acquiring the $[PSI^+]$ state from those that might be specific to individual genomes. In each genetic background we created fresh, stable isogenic $[PSI^+]$ and $[psi^-]$ pairs by guanidine hydrochloride curing²⁴. This procedure ensured that phenotypic differences detected would be due to the direct effects of $[PSI^+]$ on gene expression and physiology, rather than to secondary genetic changes that might have occurred during long-term culture. Each strain had a different intrinsic level of $[PSI^+]$ -mediated nonsense suppression (data not shown), as measured with a series of β -galactosidase test constructs⁹. Suppression levels ranged from a low of 0.2, 0.6 and 0.8% (in strain D1142-1A) to a high of 1.7, 4.5 and 16% (in strain 5V-H19) for the UAA, UAG and UGA codons, respectively in the context of this reporter gene.

We also examined NM deletion strains in two backgrounds (74D-694 and 33G-D373). These strains are deleted for the entire N and M regions of Sup35. The NM region of Sup35 might have a function other than prion maintenance, and this function might be inactivated by structural occlusion when the protein acquires the $[PSI^+]$ state. NM deletion mutants (Δ NM) would exhibit phenotypes associated with the loss of this putative function but not phenotypes due to $[PSI^+]$, because they are obligately $[psi^-]$ ¹⁸.

Yeast cells undergo many physiological changes in the transition between growth stages. Extreme care was taken to ensure that each strain pair was examined at the same growth stage. Mid-log phase cells grown in rich medium (YPD; ref. 25) were serially diluted in

Table 1 Phenotypic testing of [*PSI*⁺] / [*psi*⁻] isogenic strain pairs

	74D+	74D-	33G+	33G-	SL+	SL-	D11+	D11-	5V+	5V-	10B+	10B-	Bsc+	Bsc-
Carbon sources														
Fermentable														
Dextrose 2% (YPD) pH 6.8	5R	5R	5R	5R	5R	5R	5R	5R	5R	5R	5R	5R	5R	5R
Dextrose (YPD) pH 6	5R	5R	4M	4M	5R	5R	1S	4M	5R	4M	5R	5R	5R	5R
Dextrose (YPD) pH 4.9	4S	4S ^c	1S	1S ^c	4R	4R	NG	NG	5M	5M	5M	5R	4M	4M
Galactose 2%	5M	5M ^c	4S	4S ^c	5M	5M	5R	5R	5S	NG	5R	5R	5R	5R
Lactate 2%	5M	5M	2M	4M	5R	5R	NG	NG	5R	5R	5M	5M	5R	5R
Nonfermentable														
K-Acetate 3%	5R	5R	5S	5M ¹	5R	5R	4S	4S	5R	5R	5M	5R	5R	5R
Ethanol 3%	5M	5M	2M	4M ^c	5R	5R	5R	5S	5R	5R	5M	5M	5R	5R
Nitrogen sources														
allantoin 1 mg ml ⁻¹	NG	NG	NG	NG	2S	1S	NG	NG	2S	2S	1V	3S	NG	NG
ammonia 1 mg ml ⁻¹	NG	NG	NG	NG	2S	1S	NG	NG	1S	2S	1S	3S	NG	NG
glutamate 1 mg ml ⁻¹	4V	4V ^c	NG	NG	5V	5V	NG	NG	1S	5S	2S	5S	NG	NG
glutamine 1 mg ml ⁻¹	1V	1V ^c	NG	NG	4S	4S	NG	NG	NG	5V	5V	5S	NG	NG
ornithine 1 mg ml ⁻¹	4S	3S ^c	NG	NG	5S	5S	NG	NG	1V	5S	5V	5S	NG	NG
proline 1 mg ml ⁻¹	2S	1S ^c	NG	NG	4S	4S	NG	NG	NG	5S	2V	5S	NG	NG
serine 1 mg ml ⁻¹	2V	1V ^c	NG	NG	3S	2S	NG	NG	NG	2S	2V	5S	NG	NG
threonine 1 mg ml ⁻¹	2S	1S ^c	NG	NG	3S	3S	NG	NG	NG	5S	5V	5S	NG	NG
Salts and metals														
BaCl ₂ 50 mM	5R	5R	NG	NG	5R	5R	5M	5R	5M	5R	5R	5R	5R	5R
CaCl ₂ 0.5 M	5M	5M	5M	5M	5M	5M	NG	NG	5V	NG	5M	5R	5M	5M
CdCl ₂ 20 μM	2S	2M ^c	4S	4M ^a	2M	5M	5M	3S	5M	5R	2V	4S	NG	NG
CoCl ₂ 750 μM	5R	5R	5R	5R	5R	5R	5R	5R	5R	5R	5R	5R	5R	5R
CsCl 25 mM	5R	5R	4M	4M ^c	5M	5M	3V	4M	5M	5R	5R	5R	5R	5R
CsCl 0.1 M	NG	NG	NG	NG	5R	5R	NG	NG	NG	NG	NG	NG	5R	2R
CuSO ₄ 2.5 mM	3M	3M	5M	5R	4M	4M	4M	4R	4M	5R	5M	5R	4M	5R
CuSO ₄ 5 mM	4M	4M ^c	3S	3S ^c	5M	5M	3S	5R	5S	5R	5M	5R	5R	5R
CuSO ₄ 10 mM	NG	NG	NG	NG	NG	NG	NG	NG	NG	1S	2S	5M	3M	5M
LiCl 50 mM	5R	5R	2V	1V ^c	5R	5R	5M	5M	5S	2V	5R	5R	5R	5R
LiCl 0.15 M	3S	4R ^a	NG	NG	4M	4R	NG	NG	NG	NG	3S	2V	5M	4M
MgCl ₂ 0.5 M	4R	4R	5R	5R	4R	4R	4R	4R	5M	5R	5R	5R	5R	5R
MnCl ₂ 4 mM	5R	5R	NG	NG	5M	5R	1M	3M	5M	5M	5M	5M	4S	4S
NaCl 0.3 M	5R	5R	2S	NG ^a	4M	4M	4M	4M	4M	4V	5R	5R	5R	5R
NaCl 0.7 M	4M	4M	NG	NG	5M	5M	3M	3S	3S	3V	5M	5M	5M	5M
ZnCl ₂ 2.5 mM	5R	5R	5R	5R	5R	5R	5M	5R	5M	5R	5M	5R	5M	5R
ZnCl ₂ 5 mM	5M	5R ^a	4S ^b	4M	5M	5R	4V	5M	2V	5R	5S	5M	5M	5R
Inhibitors														
anisomycin 20 μg ml ⁻¹	1S	1S ^c	5M	2S ^c	2M	5R	NG	NG	2V	2R	3S	5M	5R	5R
benomyl 1 μg ml ⁻¹ (37 °C)	5M	5M	5M	5M	5R	5R	5R	5R	4S	3S	5M	5R	5S	5R
benomyl 20 μg ml ⁻¹	5M	5M	5M	5M	5M	5M	4S	4M	5R	5M	5M	5M	4M	4M
benomyl 20 μg ml ⁻¹ (37 °C)	5R	5R	5R	5R	5R	5R	5S	5S	4M	4S	2V	3S	1V	3V
benomyl 40 μg ml ⁻¹	4S	4S ^c	3V	5S ^c	4M	5M	1S	3S	4S	3V	3S	5M	3S	4M
bleomycin 10 μg ml ⁻¹	3S	4M ^c	3M	NG ^a	3S	4M	NG	NG	NG	3S	NG	4M	5M	5M
caffeine 10 mM	4M	4M	4R	4M	4R	4M	5R	5R	5M	3V	5R	5R	NG	NG
calcofluor white 1 mg ml ⁻¹	5S	5M ^c	4V	4M ^c	3S	4M	2M	4R	5M	5R	5M	5R	NG	NG
canavanine 30 μM (SD-arg)	5R	5R	5V	5V ^c	5R	5R	NG	NG	5M	5R	NG	NG	5S	5S
camptothecin 5 μg ml ⁻¹	4M	4M	3S	3S	4M	4M	2S	4S	5M	5M	5M	5M	5S	5S
cycloheximide 0.2 μg ml ⁻¹	2V	2V	5R	5R	1V	4S	NG	NG	NG	NG	5R	4V	3V	3S
cycloheximide 3 μg ml ⁻¹	NG	NG	NG	NG	NG	NG	NG	NG	NG	NG	4M	NG	NG	NG
diamide 1 mM	5M	5M	4M	5M ^a	5M	5M	4M	4M	4M	4M	4M	4M	4M	4M
diamide 1.5 mM	NG	NG ^c	3S	4M ^c	4M	3M	4S	4M	4S	4M	5M	5M	2S	NG
ethanol 6%	5R	5R ^c	4M	4M ^c	5R	5R	5R	3M	5R	5R	4R	4R	4R	4R
ethanol 10%	5R	5M ^a	3S	1S ^c	4R	4R	5R	5R	5M	2S	3M	3M	4R	3R
D-his 0.5 mM (L-pro)	2S	2S ^c	4S	5S ^c	5R	5R	5S	2V	5R	5R	NG	NG	4M	4M
hydroxyurea 0.1 M	5M	5M ^c	5S	4S ^c	4M	4M	4S	4S	5V	5R	5M	5R	5R	5R
neomycin 5 mg ml ⁻¹	5R	5R	4S	4S	5R	5R	3S	4M	5S	4V	5R	5R	5R	5R
4-NQO 2.5 μg ml ⁻¹	1S	1S	5M	5M	2M	3M	NG	NG	3M	4M	1V	2S	NG	NG
nystatin 2 μg ml ⁻¹	5M	5M	5S	5S	5M	5M	4V	4V	5R	5R	5R	5R	5V	5V
oligomycin 1 μg ml ⁻¹ (YPGE)	5R	5R	5R	5R	5R	5R	5R	5R	5R	5S	5R	5R	5R	5R
oligomycin 2.5 μg ml ⁻¹ (YPGE)	5V	5V	5M	5M	3V	3V	5S	5V	5M	5S	4R	4R	5M	5M
oligomycin 5 μg ml ⁻¹ (YPGE)	3V	3V	5V	5M ^a	3V	3V	5V	5V	5V	5V	4R	4R	5M	5M
paraquat 1 mM	5S	5S	5R	5M ^a	1V	4S	NG	NG	5S	5S	5S	5M	5M	5M
paraquat 5 mM	4S	1S ^a	3S	3S ^c	NG	NG	NG	NG	NG	NG	NG	4S	2S	2S
paromomycin 100 μg ml ⁻¹	5M	5M	4S	4S ^c	5M	5M	4S	4M	4S	5M	5M	5M	5M	5M
paromomycin 200 μg ml ⁻¹	5M	5M	3S	3S	5M	5M	3S	4M	4S	5M	5M	5M	5M	5M
rapamycin 0.1 μg ml ⁻¹	5M	5S ^a	1V	1V	4M	4M	NG	NG	4M	5R	5M	5R	NG	NG
staurosporine 1 μg ml ⁻¹ (37 °C)	5S	4M ^c	4V	4V ^c	5M	5M	5R	5R	5S	5S	5M	5M	1V	4S
vanadate (no KCl)	NG	NG	NG	NG	NG	NG	NG	NG	1V	2M	NG	NG	NG	NG
Miscellaneous														
K-Acetate + PB	5R	5R	3S	4M	5R	5R	5R	5R	5R	5R	5V	5R	5R	5R
Osmotic stress														
KCl 1.3 M	4M	4M	5S	5S	5S	5S	4M	4M	2S	1S	3M	4M	3M	2M
Temperature														
13 °C (YPD)	5R	5R ^c	1S	1S	4S	5R	3S	5R	5S	5R	5S	5R	5S	5R
24 °C (YPD)	5R	5R	4R	5R ^a	5R	5R	5M	5R	5R	5R	5M	5R	5M	5R
37 °C (YPD)	5R	5R	5R	5R	5R	5R	5R	5R	5R	4S	4M	5R	4R	5R
13 °C (SD)	4R	4R ^c	NG	NG	4S	5R	3S	5R	5S	5R	5S	5R	5S	5R
24 °C (SD)	5R	5R	4R	4R	5R	5R	4R	5R	5R	5R	4M	5R	5R	5R
37 °C (SD)	5R	5R	5R	5R	4R	4R	5R	5R	1M	NG	1M	2R	5R	5R
Thermotolerance	M+	M	M	M	L	L	H	H+	M	M	H	H	M	M
Ethanol gradient	H	H	M	M	H	H	M+	M	L	L	M	M	M	M

5-fold increments and spotted onto test plates. Amino-acid supplements, added in parallel experiments, did not alter growth patterns (not shown), establishing that phenotypic differences between isogenic $[PSI^+]$ and $[psi^-]$ cells were not owing to secondary effects of suppressing auxotrophic markers these strains contained. Most of the strains carried adenine mutations, some of which were suppressed by $[PSI^+]$. These produced the expected differences in red/white colony colour on YPD medium⁷. However, on this medium the growth rates of the $[PSI^+]$ derivatives in each of our isogenic pairs were always very similar to those of the corresponding $[psi^-]$ derivatives (Fig. 1, left).

In contrast, marked growth differences were observed between isogenic $[PSI^+]$ and $[psi^-]$ derivatives on many test plates. For example, when the pH of the YPD medium was changed from 6.8 to 6.0 by the addition of sodium acetate, the $[psi^-]$ derivative grew better than the $[PSI^+]$ strain in the D1142-1A background (Fig. 1, right), but in the 5V-H19 background $[PSI^+]$ cells grew better than $[psi^-]$ cells. Differences in the growth properties of isogenic $[PSI^+]$ and $[psi^-]$ cells under our test conditions are highlighted in Table 1. Representative raw data are shown in Figs 1–5. These growth differences were highly reliable. Within experiments, YPD control plates (periodically interspersed with test plates) confirmed that cells were consistently spotted at the same densities with the same efficiencies, and the same phenotypic distinctions appeared in repeat experiments (data not shown).

In nearly half of the conditions tested, $[PSI^+]$ exerted a substantial effect in at least some strains and in more than 25% of these tests, its effect was positive. The changes in growth were highly idiosyncratic (Table 1 and Fig. 2). Each strain exhibited a unique constellation of phenotypes. In some cases very modest changes in conditions produced large growth differences between isogenic $[PSI^+]$ and $[psi^-]$ cells. In others, conditions that inhibited growth affected $[PSI^+]$ and $[psi^-]$ cells similarly. Simple patterns were not readily discernible. For example, in both the 10B-H49 and 5V-H19 backgrounds, $[PSI^+]$ inhibited growth on several nitrogen sources. However, under many other test conditions the phenotypic changes induced by $[PSI^+]$ in 10B-H49 diverged from those in 5V-H19. Moreover, in the SL1010-1A and 74D-694 backgrounds, $[PSI^+]$ enhanced growth on some nitrogen sources.

In the presence of the alkali metal caesium (25 mM; Table 1, Fig. 2b), $[PSI^+]$ had little effect on most strains, but strongly inhibited growth in the D1142-1A background. In the BSC783/4c background, $[PSI^+]$ strongly enhanced growth, but only when the concentration of caesium was raised to 100 mM. In the presence of the alkali metal lithium, $[PSI^+]$ inhibited growth in 74D-694, but enhanced growth in a concentration-dependent manner in both the 5V-H19 and 10B-H49 backgrounds. In the presence of inhibitors that affect a wide variety of cellular processes^{26,27} (many of which are

naturally occurring antimicrobial agents²⁸), $[PSI^+]$ produced an equally divergent range of phenotypes (Fig. 2c, Table 1).

In some cases $[PSI^+]$ changed the rate of colony growth: on YPD, pH 6.0 (strain 5V-H19; Fig. 1) or 3% ethanol (strain D1142-1A; Fig. 2a), $[PSI^+]$ and $[psi^-]$ cells yielded similar numbers of colonies, but $[psi^-]$ colonies grew more slowly. In other cases, $[PSI^+]$ altered the extent to which cultures could be diluted and still yield viable colonies: on lactate (33G-D373) or anisomycin (SL1010-1A), $[PSI^+]$ and $[psi^-]$ colonies were almost the same size, but $[psi^-]$ cells could be diluted further and still grow. In many cases, both growth parameters were affected (Table 1 and Fig. 2). Under some conditions, $[PSI^+]$ completely inhibited growth (on bleomycin, strains 5V-H19 and 10B-H49; Fig. 2c) or enabled growth when it was otherwise inhibited (33G-D373 on bleomycin, Fig. 2c, or 5V-H19 on galactose, Fig. 2a).

Morphological changes in $[PSI^+]$ colonies

$[PSI^+]$ also produced marked changes in colony morphology. When cells were plated onto media with 3% acetate as a sole carbon source (Fig. 3) 5V-H19 $[PSI^+]$ cells produced unusually large colonies with a puckered, dimpled surface (Fig. 3a). The colony morphologies of

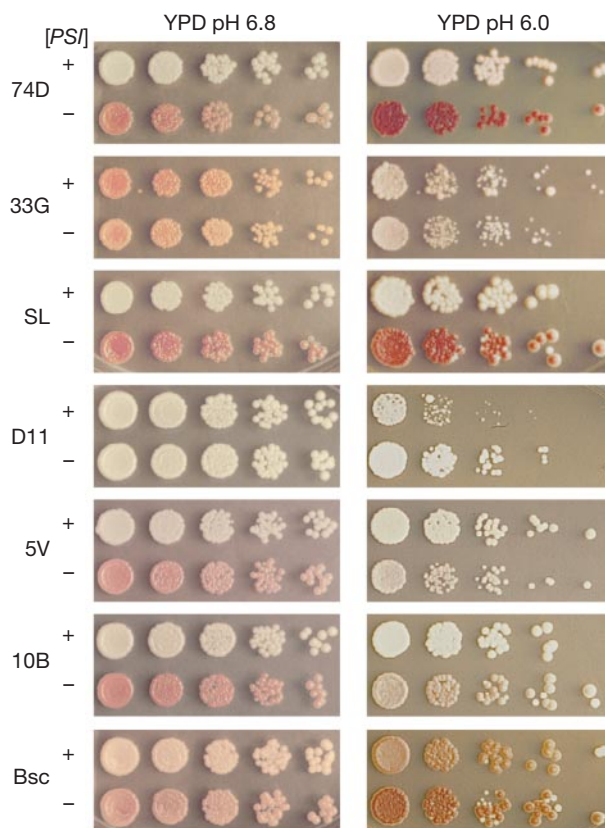


Figure 1 $[PSI^+]$ does not alter growth on standard rich media (YPD, pH 6.8), but causes strain-specific changes at pH 6.0. Exponentially growing $[PSI^+]$ and $[psi^-]$ derivatives (30 °C in YPD) of different strain backgrounds were serially diluted 5-fold, spotted on the indicated media and incubated at 30 °C. Strains: 74D, 74-D694 (*MAT* a, *ade1-14*, *trp1-289*, *his3-Δ200*, *ura3-52*, *leu2-3,112*; from Y. Chernoff⁴⁴); 33G, 33G-D373 (*MAT* a, *pheA10*, *his7-1*, *lys9-Δ21*, *trp1-289*, *ade 2-144,717*, *ura3-52*, *leu2-3,112*; from Y. Chernoff⁴⁵); SL, SL1010-1A (*MAT* α, *ade1-14*, *met 8-1*, *leu2-1*, *his5-2*, *trp1-1*, *ura 3-52*; from S. W. Liebman⁴⁶); D11, D1142-1A (*MAT* a, *aro7*, *cyc1*, *his4*, *leu2*, *lys2-187*, *met8-1*, *trp5-48*, *ura3*; from F. Sherman^{47,48}); 5V, 5V-H19 (*MAT* a, *ade2-1*, *can1-100*, *leu2-3,112*, *ura3-52*, *SUQ5*; from M. D. Ter-Avanesyan⁴⁹); 10B, 10B-H49u (*MAT* a, *ade2-1*, *lys1-1*, *his3-11,15*, *leu2-3,112*, *kar1-1*, *ura3::kanMX4*; from M. D. Ter-Avanesyan¹⁸); and BSC, BSC783/4c (*MAT* a, *ade2-1*, *SUQ5*, *leu2-3,112*, *ura3-1*, *his3-11*, *his3-15*; from M. Tuite³²). Strains 5V-H19, 10B-H49, and BSC783/4c share common ancestry as do the strains 74D-694, 33G-D373, and SL1010-1A.

Table 1 Phenotypic testing of $[PSI^+]/[psi^-]$ isogenic strain pairs. The unique pattern of phenotypes produced by $[PSI^+]$ in each strain background are designated by colour. Blue, change in colony morphology; green, enhanced growth in the $[psi^-]$ derivative; gold, enhanced growth in $[PSI^+]$, with more intense colours showing a stronger effect. Exponentially growing cells in YPD were serially diluted (5-fold from 1×10^6 cells ml^{-1}), plated and scored for growth in two ways: numbers designate the highest dilution that produced significant growth in repeat experiments, and letters indicate growth rates (R, rapid; M, medium; S, slow; V, very slow; and NG, no growth). Superscripts are used in cases where a change in growth was evident in a corresponding ΔNM strain: a indicates growth similar to $[psi^-]$; b indicates similar to $[PSI^+]$; and c indicates different from either $[PSI^+]$ or $[psi^-]$. For stress assays H, M and L are high, medium, and low growth rates, respectively, and plus indicates slightly increased tolerance. Additional conditions tested can be found in the Supplementary Information. The base medium for carbon and nitrogen source plates was prepared as described²⁷. Metals and inhibitors^{26,27} were added to YPD medium, after it was autoclaved and cooled to 55 °C. Abbreviations and additives: AntA, antimycin A; +KCl, 0.5M KCl; PB, phloxine B. The inhibitors used target the several cellular processes^{26,28} (see also Supplementary Information). The solvents used were tested separately in YPD plating assays and had no effect on growth (data not shown).

5V-H19 diploids were less extreme than those of haploids (not shown), but neither haploids nor diploids of the $[psi^-]$ derivative were affected (Fig. 3b and data not shown). Notably, in this case both $[PSI^+]$ and $[psi^-]$ cells grew at similar rates with similar plating efficiencies.

Individual cells from puckered colonies had normal morphologies by light and electron microscopy and revealed normal bud-scar distributions by calcofluor white staining²⁹ (H.L.T., T. Kowal and S.L.L., data not shown). Perhaps related to this difference in colony morphology, 5V-H19 $[PSI^+]$ cells were more flocculant than $[psi^-]$ cells, forming clumps when growing in liquid media. This colony morphology phenotype provided a vivid illustration of the tendency of yeast cells to switch spontaneously between $[PSI^+]$ states during normal growth^{7,30}. Figure 3c displays a large, puckered colony that arose in a group of 5V-H19 $[psi^-]$ cells on acetate medium. Further analysis (for example, guanidine hydrochloride curing²⁴) established that this colony phenotype was due to a spontaneous conversion to the $[PSI^+]$ state.

Interaction with agents affecting translation

$[PSI^+]$ is a prion form of the translation termination factor that changes the fidelity of protein synthesis. As translational fidelity is already compromised in $[PSI^+]$ cells, other modulators of protein synthesis might be expected to show a common response to $[PSI^+]$. Indeed, growth at lower temperatures is known to increase nonsense suppression³¹ and in five of the seven genetic backgrounds, on both rich (YPD) and minimal (SC) media²⁵, $[psi^-]$ derivatives grew better than $[PSI^+]$ derivatives at 13 °C. However, at the standard temperature (30 °C) antibiotics that specifically affect translational fidelity (anisomycin, neomycin and paromomycin), $[PSI^+]$ sometimes increased and sometimes decreased growth (Fig. 2c and Table 1). Apparently, the strain-specific effects of $[PSI^+]$ on growth supersede the direct action of the translation inhibitors.

Are there common biological effects of $[PSI^+]$?

In examining reasons for the conservation of this prion, one possibility is that $[PSI^+]$ might enhance stress tolerance by increasing heat-shock protein (Hsp) synthesis³². Hsps increase survival during exposure to heat, ethanol and other environmental stresses,

with the general signal for their induction being an increase in the cellular concentration of misfolded protein^{33,34}. $[PSI^+]$ -mediated read-through of stop codons might produce misfolded proteins in sufficient quantity to trigger this induction. Supporting this hypothesis, all three previously tested $[PSI^+]$ strains exhibited greater ethanol tolerance than their isogenic $[psi^-]$ derivatives and two of the three showed greater heat tolerance³².

We examined ethanol tolerance in our seven freshly created isogenic pairs, by growing them on agar plates containing 4% dextrose as a carbon source and a gradient of ethanol (from 0 to 10%) (Fig. 4a)³². Each genetic background had a different intrinsic tolerance to ethanol, but it made little difference whether the strains were $[PSI^+]$ or $[psi^-]$. We examined heat tolerance by treating cells at 37 °C, exposing them to 50 °C for various periods, and plating to determine the number of colony forming units remaining (Fig. 4b). Again, each genetic background exhibited a different intrinsic level of tolerance. In most cases $[PSI^+]$ had no effect, but it slightly increased survival in the 74D-694 background (20 min and 30 min time points) and slightly decreased survival in the D1142-1A background (30 min and 45 min time points). Finally, the presence or absence of $[PSI^+]$ did not correlate with the rate of growth at high temperatures (37 °C, Table 1). $[PSI^+]$ might be involved in stress tolerance under natural growth conditions not mimicked by these experiments, but its effects in standard laboratory tests were neither

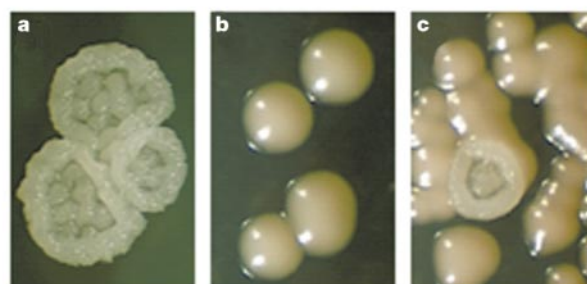


Figure 3 Colony morphology changes produced by $[PSI^+]$ in strain 5V-H19. Cells were grown in YPD medium, diluted, spotted onto medium containing potassium acetate as the sole carbon source and grown at 30 °C. **a**, $[PSI^+]$ colonies; **b**, $[psi^-]$ colonies. **c**, Spontaneous appearance of a $[PSI^+]$ colony within a group of $[psi^-]$ cells.

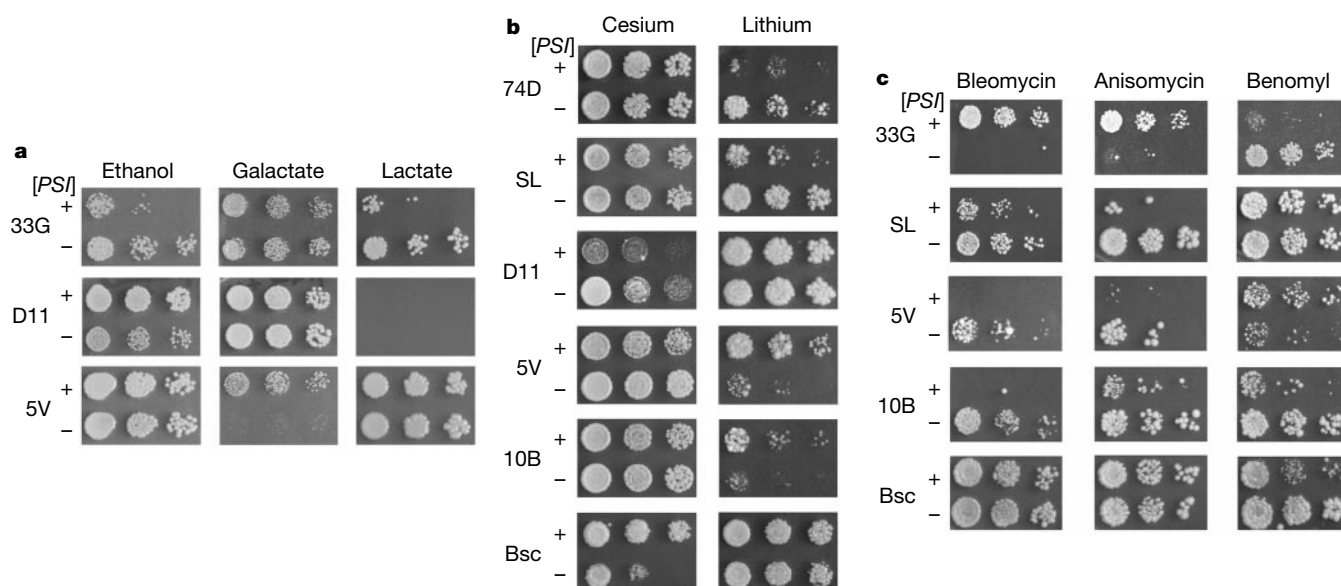


Figure 2 $[PSI^+]$ affects growth in different ways in different genetic backgrounds. **a**, Representative growth differences on media containing ethanol, galactose or lactate as the sole carbon source. **b**, Growth on media containing caesium (left) at 25 mM (100 mM for the strain BSC783/4c) or lithium (right) at 150 mM (50 mM for the strains D1142-1A

and 5V-H19). **c**, Growth on media containing the antibiotics bleomycin (10 $\mu\text{g ml}^{-1}$), anisomycin (20 $\mu\text{g ml}^{-1}$) and benomyl (40 $\mu\text{g ml}^{-1}$). Cells were grown and plated as described in Fig. 1 and Table 1.

strong nor consistent.

Of the many effects on growth we observed in response to $[PSI^+]$, in only one case did it affect all backgrounds. In all seven strain pairs, $[PSI^+]$ derivatives grew more slowly than their isogenic $[psi^-]$ derivative on media containing 5 mM $ZnCl_2$ (Fig. 5). $[PSI^+]$ also affected all six of the strains that were capable of growth in the presence of calcofluor white in the same way (Table 1). Given the number and variety of phenotypes examined, such effects might occur by chance, but it is possible that zinc metabolism or cell-wall biosynthesis reflect common physiological effects of $[PSI^+]$.

NM-specific phenotypes

Another explanation for the conservation of the prion-determining region is that it has a function separate from its role in $[PSI^+]$ formation³⁵. This possibility is supported by the fact that NM deletions produced distinct phenotypes in our assays. The most dramatic effect occurred during growth in the presence of benomyl, a microtubule-destabilizing agent. We examined benomyl toxicity extensively (Table 1 and Fig. 2c) because certain mutations in *SUP35* increase sensitivity to benomyl³⁶. Under conditions where benomyl reduced cell growth, $[PSI^+]$ generally reduced it further, but in the 5V-H19 background $[PSI^+]$ increased growth. Thus, benomyl does not reveal a common biological effect of $[PSI^+]$. However, in the two backgrounds tested, ΔNM mutants grew better in the presence of benomyl than did either of their isogenic $[PSI^+]$ or $[psi^-]$ variants. When considered with the previous data on the

SUP35 mutants³⁶, these results indicate that benomyl may perturb a function of the NM region.

Other effects of NM deletions are shown in Table 1. NM-deletion phenotypes were occasionally similar to those of the $[psi^-]$ derivative or the $[PSI^+]$ derivative, but most commonly were distinct from either. These observations support the notion that NM might play a role in yeast biology that is distinct from its role in $[PSI^+]$ formation.

Discussion

We have demonstrated that the $[PSI^+]$ element of the yeast *S. cerevisiae* provides a means to unveil silent genetic information to produce new heritable phenotypes. In the context of an individual cell, $[PSI^+]$ allows alternative heritable phenotypes to be encoded by a single genome. In the context of diverging populations it provides a vast array of new phenotypic states with unique growth advantages and disadvantages. We propose that the epigenetic and metastable nature of $[PSI^+]$, associated with the fundamental process of translation, potentiates survival in a fluctuating environment and provides a conduit for the evolution of new traits.

Yeast cells spontaneously switch from the $[psi^-]$ to the $[PSI^+]$ state³⁰, with rates generally varying between 1 in 10^5 to 1 in 10^7 . Thus, once a population has reached an appreciable size, some of its genetically identical members will have acquired a new, heritable phenotype. If the environment does not favour this phenotype, the loss of these cells will have a minimal impact on the fitness of that genotype. If the environment does favour it, the original genotype

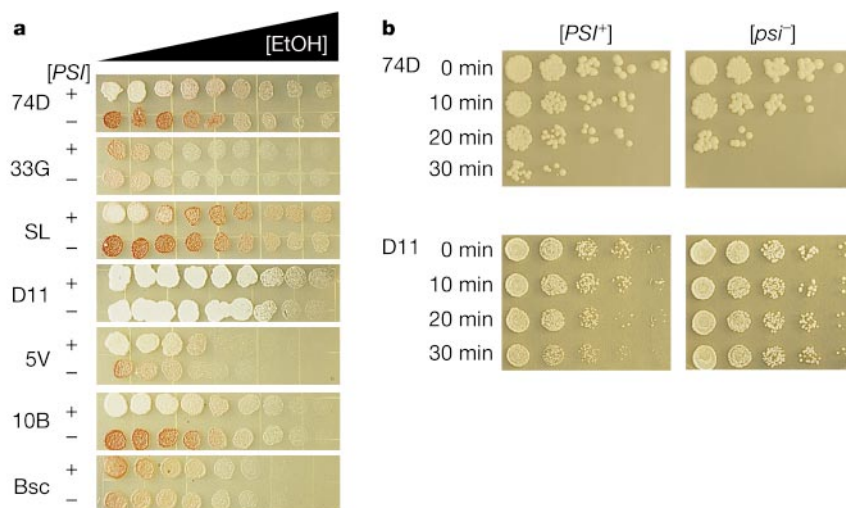


Figure 4 Stress tolerance in $[PSI^+]$ and $[psi^-]$ derivatives. **a**, For ethanol tolerance, cells were spotted at equal densities (1×10^6 cells ml^{-1}) on plates containing a gradient of ethanol from 0 to 10% in 0.25 $\times$ YEPD media³². **b**, For induced thermotolerance, cells in log phase (4×10^6 cells ml^{-1}) were incubated at 37 °C for 30 min and transferred to 50 °C

for the indicated times. Cells were serially diluted 5-fold in YPD, plated on YPD, and grown at 30 °C. Heat-shock inhibited colour development in 74D-694, but simultaneous plating to media lacking adenine demonstrated that most cells remained in the $[PSI^+]$ state.

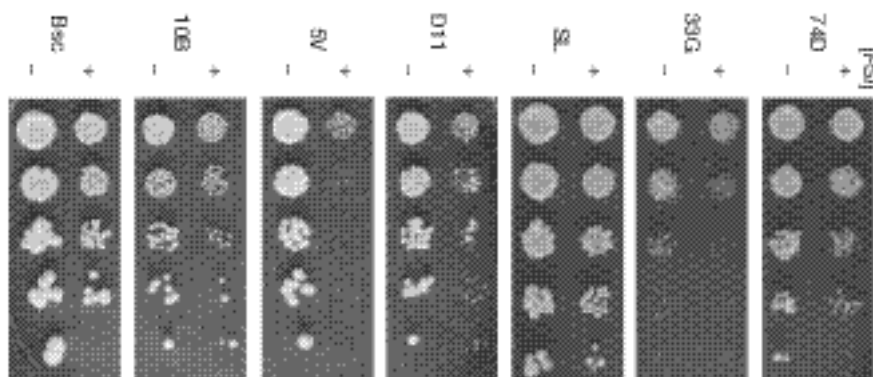


Figure 5 $[PSI^+]$ negatively affects the growth of all seven strain backgrounds in the presence of 5mM $ZnCl_2$. Cells were grown and plated as described in Table 1.

can thrive even if the original phenotype is disfavoured. Once progeny with the new phenotype have reached sufficient densities, the spontaneous appearance of $[psi^-]$ derivatives will ensure that the original phenotype is available should the environment change.

$[PSI^+]$ allows cells to occupy a new niche without foregoing their capacity to occupy the old. If the new phenotype remains advantageous, the size of the growing population increases the likelihood that mutations will arise to eliminate stop codons that are relevant to the phenotype and require $[PSI^+]$ for read-through. These would make that phenotype more robust and fix the trait; cells would retain the phenotype upon reversion to the $[psi^-]$ state (presumably normal translation fidelity will generally be favoured in nature). By this mechanism, $[PSI^+]$ could facilitate the evolution of new traits.

How might such a system evolve and be maintained? We suggest three different, not exclusive possibilities. First, the self-perpetuating conformational change that drives $[PSI^+]$ inheritance might have arisen as an inadvertent consequence of some aspect of the NM sequence that is required for another function of this region. Our data are consistent with a separate NM function, as are recent findings that NM interacts with Sla1, an actin-binding protein³⁵. Second, yeast might so commonly occupy fluctuating environments (eg. warm, nutrient-replete summers versus cool, nutrient-poor winters) that $[PSI^+]$ could be maintained as a means of alternating between heritable phenotypic states adapted to these fluctuations. Finally, the natural environment occupied by *S. cerevisiae* during its evolution may have been sufficiently erratic to provide the pressure required to maintain a global mechanism for exploiting genome-wide variation to produce new phenotypes.

$[PSI^+]$ is a suppressor of nonsense codon mutations in genetic markers³⁷. Presumably, the wealth of new phenotypes we observe in $[PSI^+]$ strains results from the read-through of two classes of naturally occurring nonsense codons: those located in open reading frames (ORFs) that have acquired inactivating stop-codon mutations (ISCs) and those employed as natural translation-termination signals at the 3' ends of active genes. ORFs might accumulate ISCs when gene duplications eliminate selective constraints (~16% of the *S. cerevisiae* proteome is present in duplicate³⁸), or when growth conditions alter selective pressures³⁹. The phenotypic diversity generated from these ORFs by $[PSI^+]$ would depend upon which ORFs had acquired ISCs, the frequency of their read-through, and the presence of additional mutations they may have acquired while in the inactive state. Any phenotypic diversity generated from the 3' untranslated regions would depend upon the context of the stop codon and the coding potential of the downstream sequence. The 3' untranslated regions of most genes are naturally under less constraint than coding sequences and are highly polymorphic⁴⁰. $[PSI^+]$ -mediated read-through might alter the stability of the messenger RNA, its capacity to be translated, and the function or stability of the encoded protein.

Many read-through events may have little phenotypic consequence, and others may have strong effects. In either case, the interplay between nonsense-mediated decay and translation termination⁴¹ might increase accumulation of the read-through transcript. Feedback regulation by the gene product might reduce it. Each primary event may produce many secondary and tertiary changes in gene expression and some, but not all, of these may contribute to phenotypic change. Preliminary DNA microarray analysis of $[PSI^+]$ -induced changes in the mRNA populations of the seven strain pairs (H.L.T., V. Iyer, P. Brown and S.L.L., unpublished observations) together with the existence of many potential ORFs with single nonsense codons (H.L.T., J. Henikoff, S. Henikoff and S.L.L., unpublished observations) in the only *S. cerevisiae* strain whose genome has been sequenced (<http://genome-www.stanford.edu/Saccharomyces/>)⁴² indicate it will not be trivial to pinpoint the causes of $[PSI^+]$ -induced phenotypic variation. However, continuing advances in genomics and proteomics should make such an understanding attainable.

It may also be possible directly to test the effects of $[PSI^+]$ on the evolution of phenotypic and genetic diversity. We obtained $[PSI^+]$ and $[psi^-]$ strain pairs from several laboratories, and they were assumed to be isogenic. However, when the $[PSI^+]$ derivatives were cured of the prion (converted to $[psi^-]$) they were phenotypically distinct from the original $[psi^-]$ derivatives. It may be that independent errors in strain husbandry occurred in each laboratory. It seems more likely that selective pressures from $[PSI^+]$ -induced changes in gene expression led, during the years the strains were cultured, to the selection of genetic variants.

We previously proposed another mechanism for uncovering hidden genetic variation and facilitating the process of evolutionary change in *Drosophila* and other eukaryotes⁴³. That mechanism involves environmentally induced changes in the capacity of the protein chaperone Hsp90 to maintain the activity of metastable regulatory proteins. There are many differences in the manner that Hsp90 and $[PSI^+]$ uncover variation. The former employs mendelian re-assortment of pre-existing variation and several rounds of selection for traits to become heritable and fixed. The latter involves an epigenetic change that produces heritable phenotypes in a single, readily reversible step, with additional mutations required for fixation. Both provide mechanisms for unveiling pre-existing variation in a combinatorial, genome-wide manner that converts neutral, or near-neutral, mutations to a non-neutral state. Such mechanisms may be present more broadly than previously suspected and exert an important influence on the rates and mechanisms of evolutionary change. □

Received 31 May; accepted 24 July 2000.

- Koch, A. L. Enzyme evolution. I. The importance of untranslatable intermediates. *Genetics* **72**, 297–316 (1972).
- Ohno, S. *Evolution by Gene Duplication* (Springer, New York, 1970).
- Ohta, T. Simulating evolution by gene duplication. *Genetics* **115**, 207–213 (1987).
- Walsh, J. B. How often do duplicated genes evolve new functions? *Genetics* **139**, 421–428 (1995).
- Nadeau, J. H. & Sankoff, D. Comparable rates of gene loss and functional divergence after genome duplications in vertebrate evolution. *Genetics* **147**, 1259–1266 (1997).
- Hall, B. G., Yokoyama, S. & Calhoun, D. H. Role of cryptic genes in microbial evolution. *Mol. Biol. Evol.* **1**, 109–124 (1983).
- Cox, B. $[PSI]$, a cytoplasmic suppressor of super-suppression in yeast. *Heredity* **20**, 505–521 (1965).
- Liebm, S. W. & Sherman, F. Extrachromosomal psi^+ determinant suppresses nonsense mutations in yeast. *J. Bact.* **139**, 1068–1071 (1979).
- Firoozan, M., Grant, C. M., Duarte, J. A. & Tuite, M. F. Quantitation of readthrough of termination codons in yeast using a novel gene fusion assay. *Yeast* **7**, 173–183 (1991).
- Serio, T. R. & Lindquist, S. L. $[PSI^+]$: an epigenetic modulator of translation termination efficiency. *Annu. Rev. Cell Dev. Biol.* **15**, 661–703 (1999).
- Stansfield, I. et al. The products of the SUP45 (eRF1) and SUP35 genes interact to mediate translation termination in *Saccharomyces cerevisiae*. *EMBO J.* **14**, 4365–4373 (1995).
- Zhouravleva, G. et al. Termination of translation in eukaryotes is governed by two interacting polypeptide chain release factors, eRF1 and eRF3. *EMBO J.* **14**, 4065–4072 (1995).
- Patino, M. M., Liu, J. J., Glover, J. R. & Lindquist, S. Support for the prion hypothesis for inheritance of a phenotypic trait in yeast. *Science* **273**, 622–626 (1996).
- Paushkin, S. V., Kushnirov, V. V., Smirnov, V. N. & Ter-Avanesyan, M. D. Propagation of the yeast prion-like $[psi^+]$ determinant is mediated by oligomerization of the SUP35-encoded polypeptide chain release factor. *EMBO J.* **15**, 3127–3134 (1996).
- Kikuchi, Y. & Kikuchi, A. in *Gene Expression and Regulation: the Legacy of Luigi Gorini* (eds Bissell, M., Deho, G., Sironi, G. & Torriani, A.) 257–264 (Elsevier, Amsterdam, 1988).
- Kushnirov, V. V. et al. Nucleotide sequence of the SUP2 (SUP35) gene of *Saccharomyces cerevisiae*. *Gene* **66**, 45–54 (1988).
- Wilson, P. G. & Culbertson, M. R. SUP12 suppressor protein of yeast. A fusion protein related to the EF-1 family of elongation factors. *J. Mol. Biol.* **199**, 559–573 (1988).
- Ter-Avanesyan, M. D., Dagkesamanskaya, A. R., Kushnirov, V. V. & Smirnov, V. N. The SUP35 omnipotent suppressor gene is involved in the maintenance of the non-Mendelian determinant $[psi^+]$ in the yeast *Saccharomyces cerevisiae*. *Genetics* **137**, 671–676 (1994).
- Liu, J. J. & Lindquist, S. Oligopeptide-repeat expansions modulate 'protein-only' inheritance in yeast. *Nature* **400**, 573–576 (1999).
- Kushnirov, V. V. et al. Divergence and conservation of SUP2 (SUP35) gene of yeast *Pichia pinus* and *Saccharomyces cerevisiae*. *Yeast* **6**, 461–472 (1990).
- Chernoff, Y. et al. Evolutionary conservation of prion-forming abilities of the yeast Sup35 protein. *Mol. Microbiol.* **35**, 865–876 (2000).
- Santoso, A., Chien, P., Osherovich, L. Z. & Weissman, J. S. Molecular basis of a yeast prion species barrier. *Cell* **100**, 277–288 (2000).
- Kushnirov, V. V., Kochneva-Pervukhova, N. V., Chechenova, M. B., Frolova, N. S. & Ter-Avanesyan, M. D. Prion properties of the Sup35 protein of yeast *Pichia methanolicus*. *EMBO J.* **19**, 324–331 (2000).
- Tuite, M. F., Mundy, C. R. & Cox, B. S. Agents that cause a high frequency of genetic change from $[psi^+]$ to $[psi^-]$ in *Saccharomyces cerevisiae*. *Genetics* **98**, 691–711 (1981).
- Adams, A., Gottschling, D. E., Kaiser, C. A., Stearns, T. *Methods in Yeast Genetics* (Cold Spring Harbor Laboratory Press, Plainview, 1997).

26. Hampsey, M. A review of phenotypes in *Saccharomyces cerevisiae*. *Yeast* **13**, 1099–1133 (1997).
27. Rieger, K. J. *et al.* Large-scale phenotypic analysis—the pilot project on yeast chromosome III. *Yeast* **13**, 1547–1562 (1997).
28. Budavari, S. (ed.) *The Merck Index* (Merck and Company, Rahway, New Jersey, 1989).
29. Roncero, C., Valdivieso, M. H., Ribas, J. C. & Duran, A. Isolation and characterization of *Saccharomyces cerevisiae* mutants resistant to Calcofluor white. *J. Bacteriol.* **170**, 1950–1954 (1988).
30. Lund, P. M. & Cox, B. S. Reversion analysis of [psi⁻] mutations in *Saccharomyces cerevisiae*. *Genet. Res.* **37**, 173–182 (1981).
31. Derkatch, I. L., Bradley, M. E., Zhou, P., Chernoff, Y. O. & Liebman, S. W. Genetic and environmental factors affecting the de novo appearance of the [PSI⁺] prion in *Saccharomyces cerevisiae*. *Genetics* **147**, 507–519 (1997).
32. Eaglestone, S. S., Cox, B. S. & Tuite, M. F. Translation termination efficiency can be regulated in *Saccharomyces cerevisiae* by environmental stress through a prion-mediated mechanism. *EMBO J.* **18**, 1974–1981 (1999).
33. Grant, C. M., Firoozan, M. & Tuite, M. F. Mistranslation induces the heat-shock response in the yeast *Saccharomyces cerevisiae*. *Mol. Microbiol.* **3**, 215–220 (1989).
34. Parsell, D. A. & Lindquist, S. The function of heat-shock proteins in stress tolerance: degradation and reactivation of damaged proteins. *Annu. Rev. Genet.* **27**, 437–496 (1993).
35. Bailleul, P. A., Newnam, G. P., Steenbergen, J. N. & Chernoff, Y. O. Genetic study of interactions between the cytoskeletal assembly protein sla1 and prion-forming domain of the release factor sup35 (eRF3) in *Saccharomyces cerevisiae*. *Genetics* **153**, 81–94 (1999).
36. Tikhomirova, V. L. & Inge-Vechtomov, S. G. Sensitivity of sup35 and sup45 suppressor mutants in *Saccharomyces cerevisiae* to the anti-microtubule drug benomyl. *Curr. Genet.* **30**, 44–49 (1996).
37. Liebman, S. W. & Derkatch, I. L. The yeast [PSI⁺] prion: making sense of nonsense. *J. Biol. Chem.* **274**, 1181–1184 (1999).
38. Seoighe, C. & Wolfe, K. H. Updated map of duplicated regions in the yeast genome. *Gene* **238**, 253–261 (1999).
39. Olson, M. V. When less is more: gene loss as an engine of evolutionary change. *Am. J. Hum. Genet.* **64**, 18–23 (1999).
40. Levitt, R. C. Polymorphisms in the transcribed 3' untranslated region of eukaryotic genes. *Genomics* **11**, 484–489 (1991).
41. Czaplinski, K. *et al.* The surveillance complex interacts with the translation release factors to enhance termination and degrade aberrant mRNAs. *Genes Dev.* **12**, 1665–1677 (1998).
42. Mewes, H. W. *et al.* Overview of the yeast genome. *Nature* **387**, 7–65 (1997).
43. Rutherford, S. L. & Lindquist, S. Hsp90 as a capacitor for morphological evolution. *Nature* **396**, 336–342 (1998).
44. Chernoff, Y. O., Lindquist, S. L., Ono, B., Inge-Vechtomov, S. G. & Liebman, S. W. Role of the chaperone protein Hsp104 in propagation of the yeast prion-like factor [psi⁺]. *Science* **268**, 880–884 (1995).
45. Chernoff, Y. O. *et al.* Conservative system for dosage-dependent modulation of translational fidelity in eukaryotes. *Biochimie* **74**, 455–461 (1992).
46. Zhou, P. *et al.* The yeast non-Mendelian factor [ETA⁺] is a variant of [PSI⁺], a prion-like form of release factor eRF3. *EMBO J.* **18**, 1182–1191 (1999).
47. Eustice, D. C., Wakem, L. P., Wilhelm, J. M. & Sherman, F. Altered 40S ribosomal subunits in omnipotent suppressors of yeast. *J. Mol. Biol.* **188**, 207–214 (1986).
48. Wakem, L. P. & Sherman, F. Isolation and characterization of omnipotent suppressors in the yeast *Saccharomyces cerevisiae*. *Genetics* **124**, 515–522 (1990).
49. Ter-Avanesyan, M. D. *et al.* Deletion analysis of the SUP35 gene of the yeast *Saccharomyces cerevisiae* reveals two non-overlapping functional regions in the encoded protein. *Mol. Microbiol.* **7**, 683–692 (1993).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank members of the Lindquist lab and M. Jensen for discussion and comments on the manuscript, and V. Iyer, P.O. Brown, J. Henikoff and S. Henikoff for collaborative efforts and for allowing us to cite unpublished work. This research was supported by the Jane Coffin Childs Memorial Fund for Medical Research (H.L.T.), the National Institutes of Health and the Howard Hughes Medical Institute.

Correspondence and requests for materials should be addressed to S.L.L. (e-mail: s-lindquist@uchicago.edu).

Direct detection of pulsations of the Cepheid star ζ Gem and an independent calibration of the period–luminosity relation

B. F. Lane*, M. J. Kuchner*, A. F. Boden†, M. Creech-Eakman‡ & S. R. Kulkarni*

* Palomar Observatory 105-24, † Infrared Processing and Analysis Center, ‡ Jet Propulsion Laboratory 171-113, California Institute of Technology, Pasadena

Cepheids are a class of variable (pulsating) stars whose absolute luminosities are related in a simple manner to their pulsational periods. By measuring the period and using the ‘period–luminosity’ relationship, astronomers can use the observed visual brightness to determine the distance to the star. Because these stars are very luminous, they can be observed in other galaxies, and therefore can be used to help determine the expansion rate of the Universe¹ (the Hubble constant). Calibration of the period–luminosity relation is a necessary first step, but the small number of sufficiently nearby Cepheids has forced the use of a number of indirect means, with associated systematic uncertainties. Here we present a distance to the Cepheid ζ Geminorum, determined using a direct measurement (by an optical interferometer) of its changes in diameter as it pulsates. Within our uncertainty of 15 per cent, our distance is in agreement with previous indirect determinations. Planned improvements to the instrument will allow us to calibrate directly the period–luminosity relation to better than a few per cent.

The Baade–Wesselink (BW) method of determining distance to a pulsating star² has long been regarded as a promising approach to calibrating the Cepheid distance scale^{3,4}. The essence of the BW method is to combine measurements of physical radius changes with measurements of angular diameter changes (the distance is their ratio). Although physical radius changes are easily determined from radial velocity measurements, obtaining angular diameter measurements of distant stars is more challenging. Lacking direct measurements of the angular diameters of Cepheids, astronomers have had to infer angular diameter from multicolour photometry, with the attendant potential penalty of systematic error. Long-baseline interferometry offers the possibility of directly measuring angular diameter changes of nearby Cepheids. However, whereas the mean angular diameters of a few Cepheids have been measured

Table 2 Relevant parameters of the calibrator and check star, including spectral type apparent H-band magnitude (m_H) and angular distance from ζ Gem

Star ID	Spectral type	m_H	θ_{UD} (mas)	θ_{LD} (mas)	Angular separation (degrees)
HD49968	K5III	2.44	1.78 ± 0.1	1.90 ± 0.1	4.1
HD48450	K4III	~ 2.4	1.98 ± 0.4	2.09 ± 0.4	9.5

The limb-darkened angular diameter (θ_{LD}) of HD49968 was based on recent spectrophotometry²⁰, while θ_{LD} for HD48450 was estimated by fitting a blackbody to archived broad-band photometry. The relation between θ_{UD} and θ_{LD} is given by the approximation²¹:

$$\theta_{UD} = \theta_{LD} \sqrt{1 - \frac{A}{3} - \frac{B}{6}};$$

here A and B are quadratic limb darkening coefficients and depend on the spectral type of the source²². For ζ Gem we used a limb darkening factor of $\theta_{UD}/\theta_{LD} = 0.963 \pm 0.006$ for all pulsational phases.

by several groups^{5–7}, the detection of angular diameter changes has been elusive.

The Cepheid variable ζ Geminorum (HD52973) is a bright ($m_V = 3.62$ – 4.18) galactic Cepheid with a 10.15-day pulsational period⁸. From previous work^{9,10} ζ Gem was expected to have an angular diameter in the range 1.2–2.2 milliarcseconds (mas; $1 \text{ mas} = 4.85 \times 10^{-9}$ radians), resolvable by the Palomar Testbed Interferometer¹¹ (PTI). We observed ζ Gem in the H band ($1.6 \mu\text{m}$) over a two-week period; see Table 1 for a log of the observations. The system response was monitored and calibrated by interleaving observations of ζ Gem, a calibration star (HD49968) and a check star (HD48450). These two stars were chosen to have angular sizes and brightness similar to those of the target star and to be located in close angular proximity on the sky (≤ 10 degrees); see Table 2.

The primary observable measured by PTI is the amplitude (often called the visibility) of the Young’s fringe pattern created by combining starlight collected from two 40-cm apertures separated by a 110-m interferometric baseline. These fringe visibilities are related in a simple manner to the apparent angular diameter of the observed star (Table 1). The calibration of the visibilities—an important step in interferometry—followed standard procedures discussed elsewhere^{12,13}. Results are shown in Table 2 and displayed in Fig. 1. It is clear that the angular diameter of ζ Gem is undergoing changes while the angular diameter of the check star is constant; the magnitude of the diameter changes is about 10%, consistent with theoretical expectations. We searched for periodicity in the vicinity of 10 days and found a significant peak in the power spectrum at $P \approx 10.2$ days. Monte Carlo simulations show that the probability that this peak is due to noise is less than 0.2%.

For this work, we integrated a previously published radial velocity

Table 1 Observations of ζ Gem made with the Palomar Testbed Interferometer between 3 March and 8 April 2000

Epoch (JD – 2,400,000.5)	Projected baseline (m)	Effective wavelength, λ_0 (μm)	V^2	θ_{UD} (mas)
51,605.226	103.60952	1.655	0.507 ± 0.012	1.679 ± 0.014
51,606.241	103.81611	1.646	0.502 ± 0.041	1.678 ± 0.046
51,607.280	103.94029	1.650	0.515 ± 0.050	1.651 ± 0.057
51,614.192	104.05741	1.651	0.450 ± 0.051	1.800 ± 0.059
51,615.180	104.83908	1.653	0.471 ± 0.027	1.740 ± 0.031
51,617.167	104.52214	1.653	0.539 ± 0.024	1.590 ± 0.028
51,618.143	106.38027	1.651	0.549 ± 0.006	1.538 ± 0.007
51,619.168	104.31558	1.653	0.557 ± 0.016	1.553 ± 0.018
51,620.169	104.14268	1.649	0.541 ± 0.024	1.589 ± 0.028
51,622.198	103.71849	1.651	0.498 ± 0.040	1.677 ± 0.046
51,643.161	103.56732	1.640	0.499 ± 0.010	1.666 ± 0.012

Each nightly observation consisted of 3–15 130-second integrations, interleaved with one calibration star and when possible one check star (see Table 2). Shown for each night is the mean epoch of observation, the mean projected baseline length, the SNR-weighted effective observing wavelength, the measured fringe visibility squared (V^2) and the best-fit uniform-disk diameter. For a circularly symmetric uniform intensity source the expected visibility is

$$V^2 = \left(\frac{2J_1(\pi B \theta_{UD} / \lambda_0)}{\pi B \theta_{UD} / \lambda_0} \right)^2$$

where J_1 is the first-order Bessel function, B is the projected aperture separation, θ_{UD} is the apparent angular diameter of the star in the uniform-disk model, and λ_0 is the centre-band wavelength of the observation. The uncertainties shown are statistical, derived from the internal scatter of the nightly measurements; the large variation in uncertainties between nights reflects variable atmospheric conditions as well as nightly variations in the amount of time spent observing the source.

curve¹⁴ for ζ Gem. We then compared these radius changes to the interferometrically measured angular diameters, allowing us to solve for three unknown parameters: the mean radius of ζ Gem (R_ζ), the distance (d), and a phase shift ($\Delta\phi$). We include the phase shift because ζ Gem is known to be a somewhat imperfect pulsator with long-term changes in its periodicity⁸. The systematic uncertainties in the fit are from three sources. (1) The largest systematic effect is due to uncertainties in the diameter of the calibration star (Table 2). (2) A second source of systematic uncertainty comes from the fact that the observed spectrum is composed of emission from the entire disk while the radial velocity varies across the disk, being zero at the limbs. Thus we need a projection factor (p-factor) to relate the observed radial velocity to changes in the physical size of the star. We used a constant p-factor of 1.36 ± 0.05 appropriate for the published radial velocity data¹⁵ with the uncertainty covering the range of plausible values. (3) Finally, there are uncertainties in the limb-darkening factors relating visibility to angular diameter.

All of the above uncertainties were propagated in the fit, resulting in separate statistical and systematic uncertainties. A fit of this model is shown in Fig. 2. The best-fit model parameters are: $d = 336 \pm 44$ (29/33) pc, $R_\zeta = 62 \pm 11$ (5/10) $R_\odot$, where $R_\odot$ is the radius of the Sun, $\Delta\phi = 0.0054 \pm 0.018$ (0.018, 0.0006) cycles, consistent with zero. Errors are given as $\sigma_{\text{total}}(\sigma_{\text{statistical}}/\sigma_{\text{systematic}})$ with $\sigma_{\text{total}}^2 = \sigma_{\text{statistical}}^2 + \sigma_{\text{systematic}}^2$. We obtained $\chi^2 = 10.0$ for 8 degrees of freedom (d.f.). We also fitted the model using uniform data weighting; the results changed by less than 1σ .

The resulting angular diameter and distance determination of ζ Gem is consistent with previous direct (but limited) measurements. Lunar occultation experiments¹⁶ have found $\theta_{\text{UD}} = 1.81 \pm 0.31$ mas at phase $\phi = 0.5961$; our best fit gives $\theta_{\text{UD}} = 1.6 \pm 0.3$ at the same phase. Our distance determination is consistent that of ref. 10, that is, a distance to ζ Gem of 358 ± 113 pc. Observations of ζ Gem have also been made using the Navy Prototype

Optical Interferometer (NPOI)⁷; these observations found a limb-darkened diameter, averaged over four epochs of observation, of 1.55 ± 0.09 mas, but did not detect pulsations. Using our best-fit parameters we have estimated the angular diameter at the same four epochs and obtain a mean limb-darkened diameter of 1.62 ± 0.3 mas (including systematic uncertainties)—in agreement with the NPOI measurement.

We note that further improvement in measurement accuracy is possible. At present the largest source of systematic uncertainty comes from the uncertainty in the angular diameter of the calibrator stars (Table 2). This uncertainty is most easily reduced by using multiple calibrators, particularly unresolved main-sequence stars. We plan future observations to reduce this uncertainty and hope to achieve 5% accuracy.

At the present level of accuracy, our results of distance and diameter are in excellent agreement with previously published results^{17,18} based on application of the BW method and multicolour photometry which yield 410 ± 75 pc and a radius of $65R_\odot$. With continuing observations, interferometry will be in a position to either empirically validate the indirect methods or identify discrepancies and thus motivate improvements in theoretical modelling. In this context, it is worth noting that our second largest source of systematic uncertainty (a factor of about 3 smaller than the uncertainty due to the diameter of the calibrator stars, discussed above) comes from uncertainties in the theoretical modelling (such as the limb-darkening and the p-factor).

Whereas we can observe only half a dozen targets with the 100-m baseline PTI, a much larger number of Cepheids are accessible to longer baseline interferometers such as NPOI and CHARA¹⁹. Thus we anticipate that over the next few years distances to several dozen Cepheids will be determined with an accuracy of a few per cent, providing a direct calibration of the Cepheid period–luminosity relation. $\square$

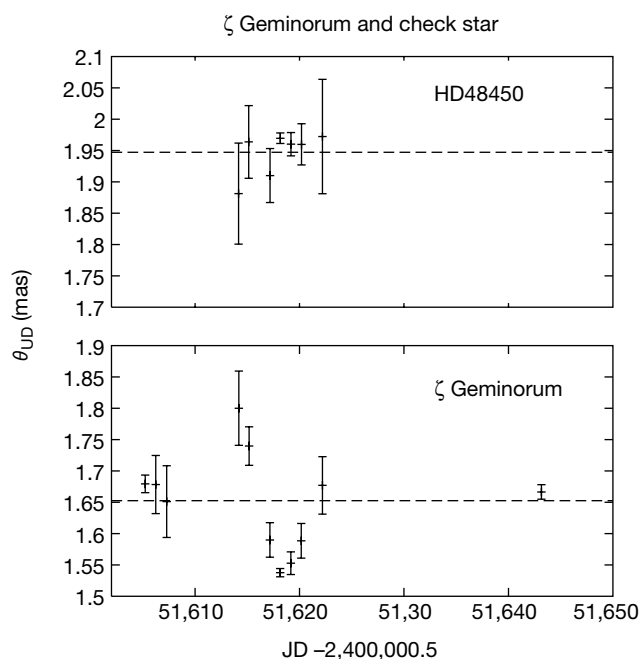


Figure 1 The uniform-disk (UD) angular diameter of ζ Gem as measured by PTI. Also shown is the UD angular diameter of the check star, HD48450. The dashed line is the unweighted mean of the data. The check star is well fitted by a constant diameter model (HD48450: $\theta_{\text{UD}} = 1.965 \pm 0.006$ mas, $\chi^2_{\text{dof}} = 0.5$). This value is consistent with the expected diameter. However, ζ Gem is not well fitted by a constant diameter model ($\chi^2_{\text{dof}} = 19.7$). We argue (see text) that changes in the angular diameter of ζ Gem are due to pulsations and in Fig. 2 we present a fit to a pulsation model.

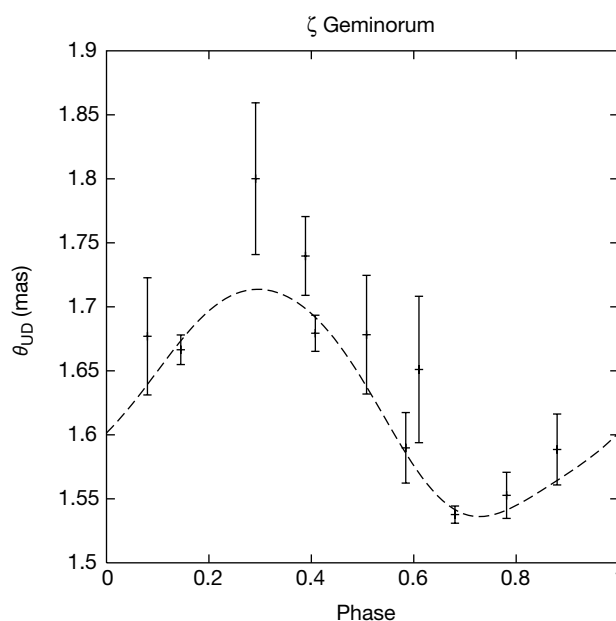


Figure 2 The UD angular diameter of ζ Gem as a function of pulsational phase with a model based on radial velocity data. Data is fitted for distance, mean radius and phase shift (see text). We assumed the period to be fixed at $P = 10.149244$ days during the time of observations, and calculated the phase shift with respect to the epoch JD 2,451,611.14505. The mean UD diameter is 1.653 mas and the real diameter is thus $1.653/0.963 = 1.717$ mas (see text and Table 2). At our inferred distance of 336 ± 44 pc we find the physical radius to be $62 \pm 11 R_\odot$.

Received 22 June; accepted 21 July 2000.

1. Mould, J. R. *et al.* The Hubble Space Telescope key project on the extragalactic distance scale. XXVIII. Combining the constraints on the Hubble constant. *Astrophys. J.* **529**, 786–794 (2000).
2. Wesselink, A. J. Surface brightnesses in the U, B, V system with applications of M_V and dimensions of stars. *Mon. Not. R. Astron. Soc.* **144**, 297–311 (1969).
3. Krokenberger, M., Sasselov, D. D. & Noyes, R. W. Radii and distances of Cepheids. I. Method and measurement errors. *Astrophys. J.* **479**, 875–885 (1997).
4. Tanvir, N. R. in *Post-Hipparcos Cosmic Candles* 17–35 (Kluwer, Dordrecht, 1999).
5. Mourard, D. *et al.* The mean angular diameter of δ Cephei measured by optical long-baseline interferometry. *Astron. Astrophys.* **317**, 789–792 (1997).
6. Kervella, P. *et al.* in *Working on the Fringe: Optical and IR Interferometry from Ground and Space* 22–27 (ASP Conf. Ser. 194, Astronomical Society of the Pacific, San Francisco, 1999).
7. Nordgren, T. *et al.* Astrophysical quantities of Cepheid variables measured with the NPOI. (in the press).
8. Szabados, L. Northern Cepheids: Period update and duplicity effects. *Commun. Konkoly Observatory, Hungary* **96**, 123–244 (1991).
9. Laney, C. D. & Stobie, R. S. The radii of galactic Cepheids. *Mon. Not. R. Astron. Soc.* **274**, 337–360 (1995).
10. Perryman, M. A. C. *et al.* The HIPPARCOS Catalogue. *Astron. Astrophys.* **323**, L49–L52 (1997).
11. Colavita, M. M. *et al.* The Palomar Testbed Interferometer. *Astrophys. J.* **510**, 505–521 (1999).
12. Colavita, M. M. Fringe visibility estimators for the Palomar Testbed Interferometer. *Publ. Astron. Soc. Pacif.* **111**, 111–117 (1999).
13. Boden, A. F. *et al.* An interferometric search for bright companions to 51 Pegasi. *Astrophys. J.* **504**, L39–L42 (1998).
14. Bersier, D., Burki, G., Mayor, M. & Duquennoy, A. Fundamental parameters of Cepheids. II. Radial velocity data. *Astron. Astrophys.* **108**, 25–39 (1994).
15. Hindsley, R. & Bell, R. A. An investigation of photoelectric radial-velocity spectrometers as used in the analysis of Cepheid variables. *Publ. Astron. Soc. Pacif.* **98**, 881–888 (1986).
16. Ridgway, S. T. *et al.* Angular diameters by the lunar occultation technique. IV—Alpha Leo and the Cepheid Zeta Gem. *Astron. J.* **87**, 680–684 (1982).
17. Gieren, W. P., Barnes, T. G. & Moffett, T. J. The period-radius relation for classical Cepheids from the visual surface brightness technique. *Astrophys. J.* **342**, 467–475 (1989).
18. Gieren, W. P., Barnes, T. G. & Moffett, T. J. The Cepheid period-luminosity relation from independent distances of 100 galactic variables. *Astrophys. J.* **418**, 135–146 (1993).
19. McAlister, H. A. *et al.* Progress on the CHARA array. *Proc. SPIE* **3350**, 947–950 (1998).
20. Cohen, M. *et al.* Spectral irradiance calibration in the infrared. X. A self-consistent radiometric all-sky network of absolutely calibrated stellar spectra. *Astron. J.* **117**, 1864–1889 (1999).
21. Welch, D. L. Near-infrared variant of the Barnes-Evans method for finding Cepheid distances calibrated with high-precision angular diameters. *Astron. J.* **108**, 1421–1426 (1999).
22. Claret, A., Diaz-Cordoves, J. & Gimenez, A. Linear and non-linear limb-darkening coefficients for the photometric bands R I J H K. *Astron. Astrophys.* **114**, 247–252 (1995).

Acknowledgements

We thank R. Akeson, T. Armstrong, A. Bouchez, M. Colavita, T. Nordgren, M. Nunez and D. Sasselov for valuable comments. Part of the work described in this paper was performed at the Jet Propulsion Laboratory under contract with the National Aeronautics and Space Administration. This research has made use of the Simbad database, operated at Centre de Données astronomiques de Strasbourg, Strasbourg, France. B.F.L. gratefully acknowledges the support of NASA through the Michelson fellowship programme. B.F.L. acknowledges the support of NASA and the NSF.

Correspondence and requests for materials should be addressed to B.F.L. (e-mail: bfl@astro.caltech.edu).

Simulating dynamical features of escape panic

Dirk Helbing*†, Illés Farkas‡ & Tamás Vicsek*‡

* Collegium Budapest—Institute for Advanced Study, Szentháromság u. 2, H-1014 Budapest, Hungary

† Institute for Economics and Traffic, Dresden University of Technology, D-01062 Dresden, Germany

‡ Department of Biological Physics, Eötvös University, Pázmány Péter Sétány 1A, H-1117 Budapest, Hungary

One of the most disastrous forms of collective human behaviour is the kind of crowd stampede induced by panic, often leading to fatalities as people are crushed or trampled. Sometimes this behaviour is triggered in life-threatening situations such as fires in crowded buildings^{1,2}; at other times, stampedes can arise during the rush for seats^{3,4} or seemingly without cause. Although engi-

neers are finding ways to alleviate the scale of such disasters, their frequency seems to be increasing with the number and size of mass events^{2,5}. But systematic studies of panic behaviour^{6–9} and quantitative theories capable of predicting such crowd dynamics^{5,10–12} are rare. Here we use a model of pedestrian behaviour to investigate the mechanisms of (and preconditions for) panic and jamming by uncoordinated motion in crowds. Our simulations suggest practical ways to prevent dangerous crowd pressures. Moreover, we find an optimal strategy for escape from a smoke-filled room, involving a mixture of individualistic behaviour and collective ‘herding’ instinct.

Up to now, panic as a particular form of collective behaviour occurring in situations of scarce or dwindling resources^{1,6} has been mainly studied from the perspective of social psychology^{7–9}. Panicking individuals tend to show maladaptive and relentless mass behaviour like jamming and life-threatening overcrowding^{1–4,8}, which has often been attributed to social contagion^{1,4,8} (see ref. 9 for an overview of theories). The observed jamming is a result of uncoordinated motion (‘incoordination’) and depends on the reward structure⁶.

We have studied related socio-psychological literature^{6–9}, reports in the media and available video materials (see <http://angel.elte.hu/~panic/>), empirical investigations^{1–3}, and engineering handbooks^{13,14}. The characteristic features of escape panics can be summarized as follows: (1) People move or try to move considerably faster than normal¹³. (2) Individuals start pushing, and interactions among people become physical in nature. (3) Moving and, in particular, passing of a bottleneck becomes uncoordinated⁶. (4) At exits, arching and clogging are observed¹³. (5) Jams build up⁷. (6) The physical interactions in the jammed crowd add up and cause dangerous pressures up to $4,450 \text{ N m}^{-1}$ (refs 2, 5) which can bend steel barriers or push down brick walls. (7) Escape is further slowed by fallen or injured people acting as ‘obstacles’. (8) People show a tendency towards mass behaviour, that is, to do what other people do^{1,8}. (9) Alternative exits are often overlooked or not efficiently used in escape situations^{1,2}.

These observations have encouraged us to model the collective phenomenon of escape panic in the framework of self-driven many-particle systems. Our computer simulations of the crowd dynamics of pedestrians are based on a generalized force model¹⁵, which is particularly suited to describing the fatal build up of pressure observed during panics^{2–5}. We assume a mixture of socio-psychological¹⁶ and physical forces influencing the behaviour in a crowd: each of N pedestrians i of mass m_i likes to move with a certain desired speed v_i^0 in a certain direction $\mathbf{e}_i^0$, and therefore tends to correspondingly adapt his or her actual velocity $\mathbf{v}_i$ with a certain characteristic time τ_i . Simultaneously, he or she tries to keep a velocity-dependent distance from other pedestrians j and walls W . This can be modelled by ‘interaction forces’ $\mathbf{f}_{ij}$ and $\mathbf{f}_{iW}$, respectively. In mathematical terms, the change of velocity in time t is then given by the acceleration equation

$$m_i \frac{d\mathbf{v}_i}{dt} = m_i \frac{v_i^0(t)\mathbf{e}_i^0(t) - \mathbf{v}_i(t)}{\tau_i} + \sum_{j \neq i} \mathbf{f}_{ij} + \sum_W \mathbf{f}_{iW} \quad (1)$$

while the change of position $\mathbf{r}_i(t)$ is given by the velocity $\mathbf{v}_i(t) = d\mathbf{r}_i/dt$. We describe the psychological tendency of two pedestrians i and j to stay away from each other by a repulsive interaction force $A_i \exp[(r_{ij} - d_{ij})/B_i] \mathbf{n}_{ij}$, where A_i and B_i are constants. $d_{ij} = ||\mathbf{r}_i - \mathbf{r}_j||$ denotes the distance between the pedestrians’ centres of mass, and $\mathbf{n}_{ij} = (n_{ij}^x, n_{ij}^y) = (\mathbf{r}_i - \mathbf{r}_j)/d_{ij}$ is the normalized vector pointing from pedestrian j to i . The pedestrians touch each other if their distance d_{ij} is smaller than the sum $r_{ij} = (r_i + r_j)$ of their radii r_i and r_j . In this case, we assume two additional forces inspired by granular interactions^{17,18}, which are essential for understanding the particular effects in panicking crowds: a ‘body force’ $k(r_{ij} - d_{ij}) \mathbf{n}_{ij}$ counteracting body compression and a ‘sliding friction force’ $\kappa(r_{ij} - d_{ij}) \Delta \mathbf{v}_{ij}^t$ impeding relative tangential motion, if pedes-

trian i comes close to j . Here $\mathbf{t}_{ij} = (-n_{ij}^2, n_{ij}^1)$ means the tangential direction and $\Delta v_{ij}^t = (\mathbf{v}_j - \mathbf{v}_i) \cdot \mathbf{t}_{ij}$ the tangential velocity difference, while k and κ represent large constants. In summary, we have

$$\mathbf{f}_{ij} = \{A_i \exp[(r_{ij} - d_{ij})/B_i] + kg(r_{ij} - d_{ij})\} \mathbf{n}_{ij} + \kappa g(r_{ij} - d_{ij}) \Delta v_{ij}^t \mathbf{t}_{ij} \quad (2)$$

where the function $g(x)$ is zero if the pedestrians do not touch each other ($d_{ij} > r_{ij}$), and is otherwise equal to the argument x .

The interaction with the walls is treated analogously: that is, if d_{iW} means the distance to wall W , $\mathbf{n}_{iW}$ denotes the direction perpendicular to it, and $\mathbf{t}_{iW}$ the direction tangential to it, the corresponding interaction force with the wall is given by:

$$\mathbf{f}_{iW} = \{A_i \exp[(r_i - d_{iW})/B_i] + kg(r_i - d_{iW})\} \mathbf{n}_{iW} - \kappa g(r_i - d_{iW})(\mathbf{v}_i \cdot \mathbf{t}_{iW}) \mathbf{t}_{iW} \quad (3)$$

Probably owing to the fact that escape panics are unexpected and dangerous events, which excludes real-life experiments, we could not find suitable data on escape panics to test our model quantitatively. This scarcity of data calls for reliable models. We have, therefore, specified the parameters as follows: with a mass of $m = 80$ kg, we represent an average soccer fan. The desired velocity v_i^0 can reach more than 5 m s^{-1} (up to 10 m s^{-1})¹⁴, but the observed free velocities for leaving a room correspond to $v_i^0 \approx 0.6 \text{ m s}^{-1}$ under relaxed, $v_i^0 \approx 1 \text{ m s}^{-1}$ under normal, and $v_i^0 \approx 1.5 \text{ m s}^{-1}$ under nervous conditions¹³. A reasonable estimate for the acceleration time is $\tau_i = 0.5$ s. With $A_i = 2 \times 10^3$ N and $B_i = 0.08$ m we can

reproduce the distance kept at normal desired velocities¹⁴ and fit the measured flows through bottlenecks¹⁴: specifically, 0.73 persons per second for an effectively 1-m-wide door under conditions with $v_i^0 \approx 0.8 \text{ m s}^{-1}$. The parameters $k = 1.2 \times 10^5 \text{ kg s}^{-2}$ and $\kappa = 2.4 \times 10^5 \text{ kg m}^{-1} \text{ s}^{-1}$ determine the obstruction effects in cases of physical interactions. Although, in reality, most parameters are varying individually, we chose identical values for all pedestrians to minimize the number of parameters for reasons of calibration and robustness, and to exclude irregular outflows because of parameter variations. However, to avoid model artefacts (gridlocks by exactly balanced forces in symmetrical configurations), a small amount of irregularity of almost arbitrary kind is needed. This irregularity was introduced by uniformly distributed pedestrian diameters $2r_i$ in the interval [0.5 m, 0.7 m], approximating the distribution of shoulder widths of soccer fans.

Based on the above model assumptions, we will now simulate several important phenomena of escape panic, which are insensitive to reasonable parameter variations, but fortunately become less pronounced for wider exits.

(1) Transition to incoordination due to clogging. The simulated outflow from a room is well coordinated and regular, if the desired velocities $v_i^0 = v_0$ are normal. But for desired velocities above 1.5 m s^{-1} , that is, for people in a rush, we find an irregular succession of arch-like blockings of the exit and avalanche-like bunches of leaving pedestrians when the arches break (Fig. 1a, b). This phenomenon is compatible with the empirical observations mentioned above, and comparable to intermittent clogging found

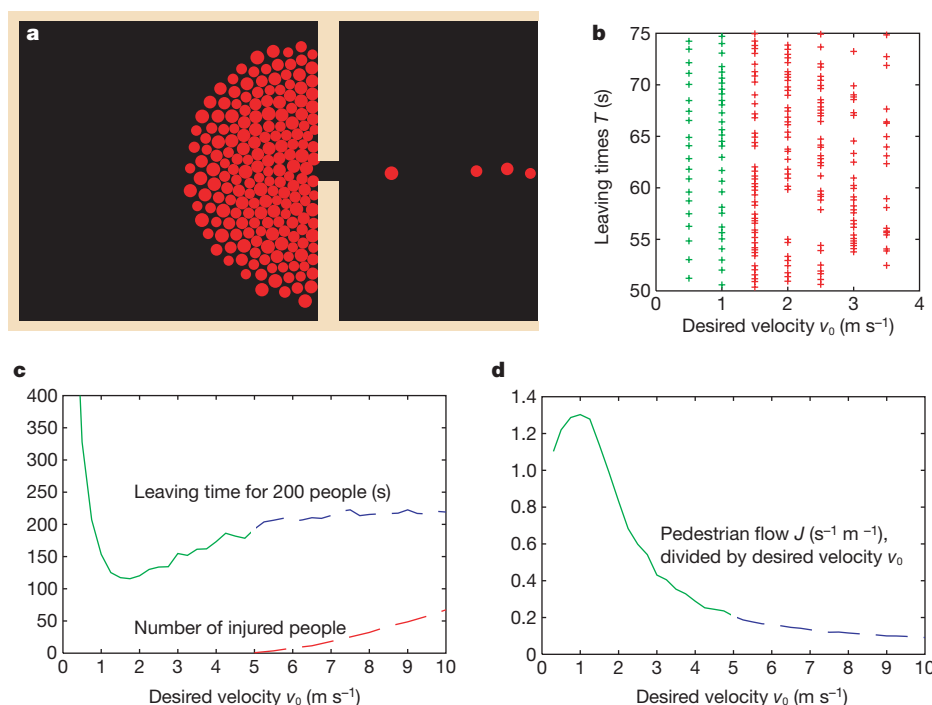


Figure 1 Simulation of pedestrians moving with identical desired velocity $v_i^0 = v_0$ towards the 1-m-wide exit of a room of size $15 \text{ m} \times 15 \text{ m}$. **a**, Snapshot of the simulation. Dynamic simulations are available at <http://angel.elte.hu/~panic/>. **b**, Leaving times of pedestrians for various desired velocities v_0 . Irregular outflow due to clogging is observed for high desired velocities ($v_0 \geq 1.5 \text{ m s}^{-1}$, red plus signs). **c**, Under conditions of normal walking, the time for 200 pedestrians to leave the room decreases with growing v_0 . Desired velocities higher than 1.5 m s^{-1} reduce the efficiency of leaving, which becomes particularly clear when the outflow J is divided by the desired velocity (**d**). This is due to pushing, which causes additional friction effects. Moreover, above a desired velocity of about $v_0 = 5 \text{ m s}^{-1}$ (corresponding to dashed lines in **c** and **d**) people are injured and become non-moving obstacles for others, if the sum of the magnitudes of the radial forces acting on them divided by their circumference exceeds a pressure of $1,600 \text{ N m}^{-1}$ (ref. 5).

Owing to the above 'faster-is-slower effect', panics can be triggered by pedestrian counterflows², which cause delays to the crowd intending to leave. This makes the stopped pedestrians impatient and pushy which may be described by increasing the desired velocity according to $v_i^0(t) = [1 - p_i(t)]v_i^0(0) + p_i(t)v_i^{\text{max}}$, where $v_i^0(0)$ is the initial, and v_i^{max} the maximum desired velocity. The time-dependent parameter $p_i(t) = 1 - \bar{v}_i(t)/v_i^0$, where $\bar{v}_i(t)$ denotes the average speed in the desired direction of motion, is a measure of impatience. Altogether, long waiting times increase the desired velocity, which can produce inefficient outflow. This further increases the waiting times, and so on, so that this tragic feedback can eventually trigger panics. It is therefore imperative to have sufficiently wide exits and to prevent counterflows when big crowds want to leave.

in granular flows through funnels or hoppers^{17,18}. (We note, however, that such clogging has been attributed to static friction between particles without remote interactions, and that the transition to clogging has been observed for small enough openings rather than for a variation of the driving force.

(2) 'Faster-is-slower effect' due to impatience. Since clogging is connected with delays, trying to move faster (that is, increasing v_i^0) can cause a smaller average speed of leaving, if the friction parameter κ is large enough (Fig. 1c, d). This effect is particularly tragic in the presence of fires, where fleeing people can reduce their own chances of survival. The related fatalities can be estimated by the number of pedestrians reached by the fire front (see <http://angel.elte.hu/~panic/>).

Because our friction term has, on average, no deceleration effect in the crowd, if the walls are sufficiently remote, the arching underlying the clogging effect requires a combination of two effects: first, slowing down due to a bottleneck such as a door, and second, strong inter-personal friction, which becomes dominant when pedestrians get too close to each other. Consequently, the danger of clogging can be minimized by avoiding bottlenecks in the construction of stadia and public buildings. We note, however, that jamming can also occur at widenings of escape routes. This surprising result is illustrated in Fig. 2. Improved outflows can be reached by columns placed asymmetrically in front of the exits, which also prevent the build up of fatal pressures (see <http://angel.elte.hu/~panic/>).

(3) Mass behaviour. We investigate a situation in which pedestrians are trying to leave a smoky room, but first have to find one of the invisible exits (see Fig. 3a). Each pedestrian i may either select an

individual direction $\mathbf{e}_i$ or follow the average direction $\langle \mathbf{e}_j^0(t) \rangle_i$ of his neighbours j in a certain radius R_i (ref. 19), or try a mixture of both. We assume that both options are weighted with some parameter p_i :

$$\mathbf{e}_i^0(t) = \text{Norm}[(1 - p_i)\mathbf{e}_i + p_i \langle \mathbf{e}_j^0(t) \rangle_i] \quad (4)$$

where $\text{Norm}(\mathbf{z}) = \mathbf{z}/\|\mathbf{z}\|$ denotes normalization of a vector $\mathbf{z}$. As a consequence, we have individualistic behaviour if p_i is low, but herding behaviour if p_i is high. Therefore, p_i reflects the degree of panic of individual i .

Our model suggests that neither individualistic nor herding behaviour performs well (see Fig. 3b). Pure individualistic behaviour means that each pedestrian finds an exit only accidentally, while pure herding behaviour implies that the entire crowd will eventually move into the same and probably blocked direction, so that available exits are not efficiently used (Fig. 3d), in agreement with observations. According to Fig. 3b and c, we expect optimal chances of survival for a certain mixture of individualistic and herding behaviour, where individualism allows some people to detect the exits and herding guarantees that successful solutions are imitated by the others. If pedestrians follow the walls instead of 'reflecting' at them, we expect that herd following also causes jamming and inefficient use of doors (see Fig. 1), while individualists moving in opposite directions obstruct each other.

Our continuous pedestrian model is based on plausible interactions, and, owing to its simplicity, is robust with respect to parameter variations. Therefore, it is suitable for drawing conclusions about the possible mechanisms underlying the effects of escape panic (regarding an increase of the desired velocity, strong friction effects during physical interactions, and herding). After

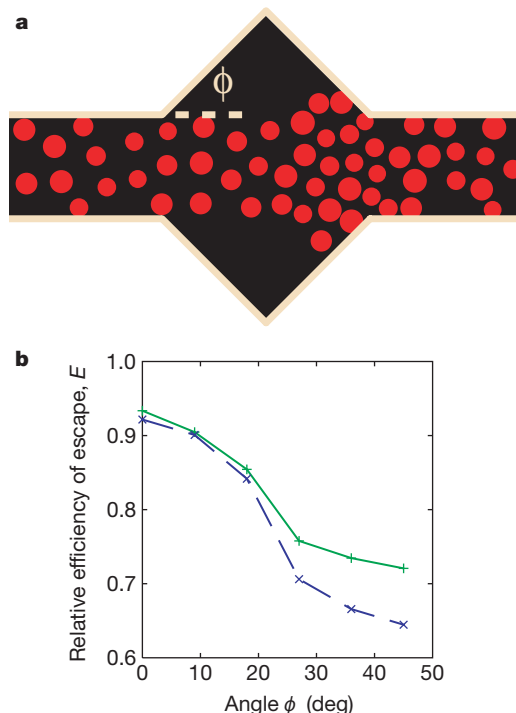


Figure 2 Simulation of an escape route with a wider area. **a**, Snapshot of the simulation with $v_i^0 = v_0 = 2 \text{ m s}^{-1}$. (Dynamic simulations are available at <http://angel.elte.hu/~panic/>) The corridor is 3 m wide and 15 m long, the length of the triangular pieces in the middle being $2 \times 3 \text{ m} = 6 \text{ m}$. Pedestrians enter the simulation area on the left-hand side with an inflow of $J = 5.5 \text{ s}^{-1} \text{ m}^{-1}$ and flee towards the right-hand side. **b**, Efficiency of leaving as a function of the angle ϕ characterizing the width of the central zone, that is, the difference from a linear corridor. The relative efficiency $E = \langle \mathbf{v} \cdot \mathbf{e}^0 \rangle / v_0$ measures the average velocity along the corridor compared to the desired velocity and lies between 0 and 1 (solid line). While it is almost one (that is, maximal) for a linear corridor ($\phi = 0$), the

efficiency drops by about 20% if the corridor contains a widening. This is understandable if we take into account that the widening leads to disturbances by pedestrians, who increase their separations in the wide area because of their repulsive interactions or try to overtake each other, and squeeze into the main stream again at the end of the widening. Hence the right half of the illustrated corridor acts like a bottleneck and leads to jamming. The drop of efficiency E is even more pronounced (1) in the area of the widening where pedestrian flow is most irregular (dashed line), (2) if the corridor is narrow, (3) if the pedestrians have different or high desired velocities, and (4) if the pedestrian density in the corridor is higher.

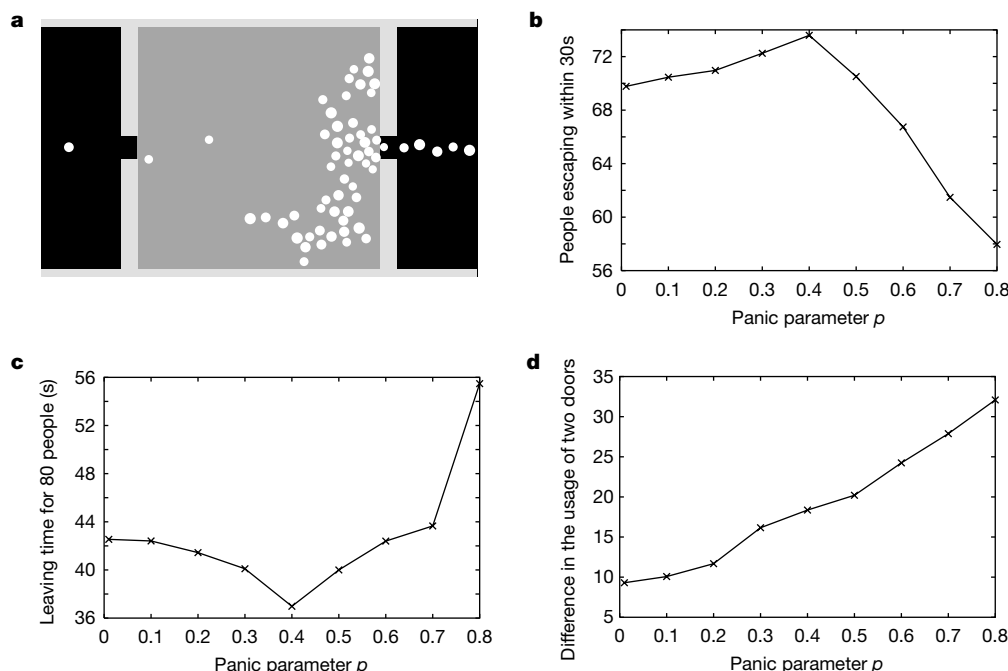


Figure 3 Simulation of $N = 90$ pedestrians trying to escape a smoky room of area $A = 15 \text{ m} \times 15 \text{ m}$ (grey) through two invisible doors of 1.5 m width. These doors have to be found with a mixture of individualistic and herding behaviour. Dynamic simulations are available at <http://angel.elte.hu/~panic/>. **a**, Snapshot of the simulation with $v_i^0 = v_0 = 5 \text{ m s}^{-1}$. Initially, each pedestrian selects his or her desired walking direction randomly. Afterwards, a pedestrian's walking direction is influenced by the average direction of the neighbours within a radius of, for example, $R_i = R = 5 \text{ m}$. The strength of this herding effect grows with increasing panic parameter $p_i = p$ and increasing value of $h = \pi R^2 \rho$, where $\rho = N/A$ denotes the pedestrian density. When reaching a boundary, the direction of a pedestrian is reflected. If one of the exits is closer than 2 m, the room is left. **b**, Number of people who manage to escape within 30 s as a function of the panic parameter p . **c**, Illustration of the time required by 80 individuals to leave the

smoky room. If the exits are relatively narrow and the panic parameter p is small or large, leaving takes a particularly long time, so that only some of the people escape before being poisoned by smoke. Our results suggest that the best escape strategy is a certain compromise between the following of others and an individualistic searching behaviour. This fits well with experimental data on the efficiency of group problem solving²⁰, according to which groups normally perform better than individuals, but masses are inefficient in finding new solutions to complex problems. **d**, Absolute difference $|N_1 - N_2|$ in the numbers N_1 and N_2 of persons leaving through the left exit or the right exit as a function of the panic parameter p . We find that pedestrians tend to jam up at one of the exits instead of equally using all available exits, if the panic parameter is large.

having calibrated the model parameters to available data on pedestrian flows, we have reproduced many observed phenomena. Our model could be used to test buildings for their suitability in emergency situations. Moreover, it accounts for the different dynamics in normal and panic situations by changing a single parameter $p_i = p$.

We are now calling for complementary data and additional video material on escape panics to test our model quantitatively, and compare it with alternative models. Such models could, for example, include direction- and velocity-dependent interpersonal interactions, specify the individual variation of parameters, study the effect of fluctuations, consider falling people, integrate acoustic information exchange, implement more complex strategies and interactions (also three-dimensional ones), or allow for switching of strategies. A superior theory would have to reproduce the empirical findings equally well with less parameters, reach a better quantitative agreement with data with the same number of parameters, or reproduce additional observations. $\square$

Received 2 June; accepted 31 August 2000.

- Keating, J. P. The myth of panic. *Fire J.* 57–61, 147 (May 1982).
- Elliot, D. & Smith, D. Football stadia disasters in the United Kingdom: learning from tragedy? *Ind. Environ. Crisis Q.* 7(3), 205–229 (1993).
- Jacobs, B. D. & 't Hart, P. in *Hazard Management and Emergency Planning* (eds Parker, D. J. & Handmer, J. W.) Ch. 10 (James & James Science, London, 1992).
- Johnson, N. R. Panic at "The Who Concert Stampede": an empirical assessment. *Soc. Problems* 34(4), 362–373 (1987).
- Smith, R. A. & Dickie, J. F. (eds) *Engineering for Crowd Safety* (Elsevier, Amsterdam, 1993).
- Mintz, A. Non-adaptive group behavior. *J. Abnormal Normal Social Psychol.* 46, 150–159 (1951).
- Kelley, H. H., Condry, J. C., Jr, Dahlke, A. E. & Hill, A. H. Collective behavior in a simulated panic

situation. *J. Exp. Social Psychol.* 1, 20–54 (1965).

- Quarantelli, E. The behavior of panic participants. *Sociol. Social Res.* 41, 187–194 (1957).
- Brown, R. *Social Psychology* (Free, New York, 1965).
- Drager, K. H. et al. in *Proc. 1992 Emergency Management and Engineering Conf.* (ed. Drager, K. H.) 101–108 (Society for Computer Simulation, Orlando, Florida, 1992).
- Ebihara, M., Ohtsuki, A. & Iwaki, H. A model for simulating human behavior during emergency evacuation based on classificatory reasoning and certainty value handling. *Microcomput. Civil Eng.* 7, 63–71 (1992).
- Still, G. K. New computer system can predict human behaviour response to building fires. *Fire* 84, 40–41 (1993).
- Predtetschenski, W. M. & Milinski, A. I. *Personenströme in Gebäuden—Berechnungsmethoden für die Projektierung* (Müller, Köln-Braunsfeld, 1971).
- Weidmann, U. *Transporttechnik der Fußgänger* (Institut für Verkehrsplanung, Transporttechnik, Straßen- und Eisenbahnbau (IVT), ETH Zürich, 1993).
- Helbing, D., Farkas, I. J. & Vicsek, T. Freezing by heating in a driven mesoscopic system. *Phys. Rev. Lett.* 84, 1240–1243 (2000).
- Helbing, D. & Molnár, P. Social force model for pedestrian dynamics. *Phys. Rev. E* 51, 4282–4286 (1995).
- Ristow, G. H. & Herrmann, H. J. Density patterns in two-dimensional hoppers. *Phys. Rev. E* 50, R5–R8 (1994).
- Wolf, D. E. & Grassberger, P. (eds) *Friction, Arching, Contact Dynamics* (World Scientific, Singapore, 1997).
- Vicsek, T., Czirók, A., Ben-Jacob, E., Cohen, I. & Schochet, O. Novel type of phase transition in a system of self-driven particles. *Phys. Rev. Lett.* 75, 1226–1229 (1995).
- Kelley, H. H. & Thibaut, J. W. in *The Handbook of Social Psychology* Vol. 4 (eds Lindzey, G. & Aronson, E.) (Addison-Wesley, Reading, Massachusetts, 1969).

Acknowledgements

D.H. thanks the German Research Foundation (DFG) for financial support by a Heisenberg scholarship. T.V. and I.F. are grateful for partial support by OTKA and FKFP.

Correspondence and requests for materials should be addressed to D.H. (e-mail: helbing@trafficforum.de).

Single photons on demand from a single molecule at room temperature

B. Lounis*† & W. E. Moerner*

* Department of Chemistry, Stanford University, Stanford, California 94305-5080, USA

† CPMOH, Université Bordeaux I, 33405 Talence, France

The generation of non-classical states of light¹ is of fundamental scientific and technological interest. For example, 'squeezed' states² enable measurements to be performed at lower noise levels than possible using classical light. Deterministic (or triggered) single-photon sources exhibit non-classical behaviour in that they emit, with a high degree of certainty, just one photon at a user-specified time. (In contrast, a classical source such as an attenuated pulsed laser emits photons according to Poisson statistics.) A deterministic source of single photons could find applications in quantum information processing³, quantum cryptography⁴ and certain quantum computation problems⁵. Here we realize a controllable source of single photons using optical pumping of a single molecule in a solid. Triggered single photons are produced at a high rate, whereas the probability of simultaneous emission of two photons is nearly zero—a useful property for secure quantum cryptography. Our approach is

characterized by simplicity, room temperature operation and improved performance compared to other triggered sources of single photons.

Various schemes have been proposed to create a single-photon source, for example, involving single atoms in cavities^{6–8}, or highly nonlinear cavities⁹. Until now, such sources have only been demonstrated in two experiments at cryogenic temperatures. The first such demonstration utilized a "turnstile" effect¹⁰ based on a Coulomb blockade for electrons and holes in a mesoscopic double-barrier p–n junction¹¹. A dilution refrigerator operating at 50 mK was essential to minimize the thermal energy background. Owing to the sample geometry, the detection efficiency was limited to about 10^{-4} , which made it difficult to measure the second-order intensity correlation of the emitted photons, an important diagnostic of the non-classical nature of the light.

A second approach is based on the controlled excitation of single molecules in a solid¹². A rapid adiabatic passage method is used to prepare the molecule in its fluorescent state. Very narrow optical absorption lines are needed for this method, which requires liquid helium temperatures (2 K). The detection efficiency was limited to about 3×10^{-3} , mainly because filtering out the pumping laser excitation also eliminates single-molecule fluorescence near the laser frequency.

Our source of single photons on demand is based on a simple concept, the pulsed optical excitation of a single highly fluorescent molecule as illustrated in Fig. 1a. A short pulse of green laser light pumps the four-level scheme of the molecule from the ground singlet state to a vibronically excited level of the first electronic excited singlet state. After fast (picoseconds) relaxation to the lowest electronic excited state, the molecule subsequently emits a single photon. Because the molecule can be pumped at high intensity, the probability of preparing the emitting state can approach unity. This scheme has the further advantage of spectrally separating the laser excitation and the fluorescence emission.

Fluorescence microscopy at ambient conditions of single emissive dye molecules^{13,14} embedded in polymers, droplets and gels or dispersed on surfaces has revealed various previously unknown effects that are normally obscured by averaging in conventional ensemble measurements. Digital photobleaching¹⁵ is an example; however, it imposes an upper limit on the number of photons detectable from a molecule. The photobleaching quantum efficiency of most single molecules has been reported to range from about 10^{-4} to about 10^{-7} in various host matrices, which is a severe limitation for quantum optical experiments^{16–18}. However,

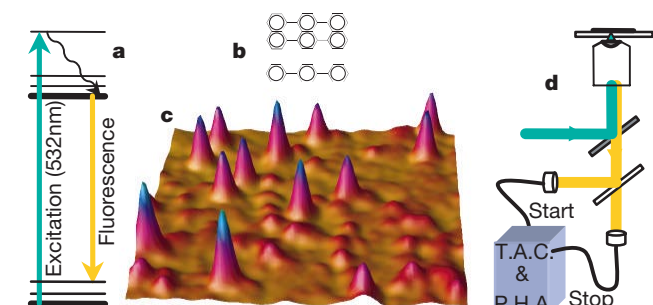


Figure 1 Scheme of the experimental measurements with images of single molecules. **a**, Pumping scheme of the molecules to their fluorescent state. **b**, Chemical structures of terrylene (upper) and p-terphenyl (lower). The sample is a sublimed crystal flake (a few micrometres in thickness) of p-terphenyl doped with terrylene at low concentrations of about 10^{-11} moles of terrylene per mole of p-terphenyl. Terrylene fluoresces at around 579 nm in the p-terphenyl crystal, and has a fluorescence quantum yield close to unity. **c**, Confocal fluorescence image ($10 \mu\text{m} \times 10 \mu\text{m}$) of single terrylene molecules embedded in crystalline p-terphenyl with continuous-wave excitation at 532 nm of about $1.5 \mu\text{W}$, with signal-to-background ratio greater than 5. **d**, Schematic representation of the scanning confocal microscope set-up used to characterize the single-molecule emission. The laser excitation was provided by a continuous-wave 532 nm frequency-doubled Nd-YAG laser for the antibunching measurement. For the pulsed excitation we used an actively mode-locked picosecond Nd-YAG laser from Lightwave Electronics (model 131, pulse width 35 ps, repetition rate 100 MHz, average power 220 mW). A pulse picker was used to reduce the repetition rate to 6.25 MHz. The pulsed green light was produced by single pass second-harmonic generation in a periodically poled lithium niobate crystal²⁸ (a maximum average green power of 0.2 mW was obtained for an average near infrared power of 10 mW). The excitation beam was directed onto the sample by the dichroic mirror of an inverted microscope and was focused on the sample plane by a 1.4 NA (numerical aperture) oil-immersion objective. The fluorescence photons were collected by this objective, and filtered from the residual excitation light by a holographic notch filter and long pass glass filter. The emitted fluorescence was sent through a 50/50 beam splitter onto two (start and stop) Si avalanche photon-counting photodiodes. The histogram of the delays between detected photons was measured using a time-to-amplitude converter and a pulse-height analyser. The stop signal was delayed for negative times.

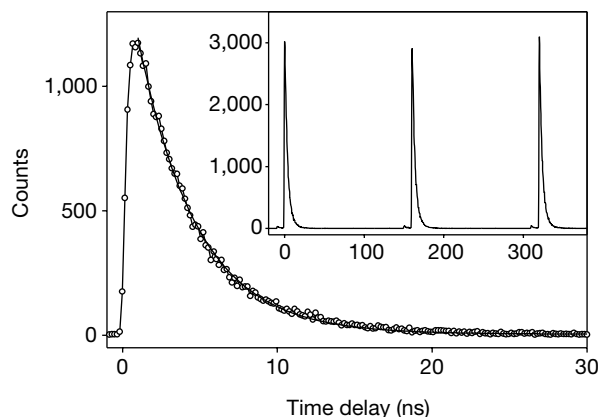


Figure 2 Characterization of the average time structure of the source. The arrival times of fluorescence photons are recorded with one detector using a time-to-amplitude converter triggered by the laser pulse. The average power of the pulsed laser is $6 \mu\text{W}$ and the acquisition time is 60 s. Circles represent the histogram of arrival times of detected photons after the pump pulse at $t = 0$. The line is a single exponential decay fit (decay time of 3.8 ns). Inset, wide time range.

highly stable single terrylene molecules have recently been detected in p-terphenyl molecular crystals at room temperature^{19,20} (for structures see Fig. 1b). In this organic crystal, the fluorescent molecules are protected by the host from exposure to diffusing quenchers (such as oxygen), and profit from the ability to emit host phonons to prevent thermally induced damage. We find that for thick crystals ($\sim 10\ \mu\text{m}$), this system indeed has extremely high photostability, allowing hours of continuous illumination without photobleaching.

Figure 1c shows a scanning confocal microscope image of the fluorescence from single terrylene molecules embedded in a thin, sublimed platelet of crystalline p-terphenyl. The isolated peaks ($\sim 400\ \text{nm}$ full-width at half-maximum) represent the emission of single molecules. Saturation studies that we performed on this system show that maximum emission rates as high as 2.5 MHz are achievable, and typical saturation intensities are close to $1\ \text{MW cm}^{-2}$. The high saturation intensities are the result of the orientation of the molecular transition dipole moment which is nearly perpendicular to the plane of the sample and hence to the laser polarization.

The light emitted by a single molecule excited by a continuous-wave laser consists of single photons separated by random time intervals which depend on the excited state lifetime and the pumping rate. Because a single molecule can never emit two photons at once, the distribution of photon pairs separated by a time τ rises from zero at $\tau = 0$. This phenomenon, called photon antibunching, has previously been observed for single molecules at low²¹ and room temperature²². Using the standard Hanbury-Brown and Twiss coincidence set-up as shown in Fig. 1d, we observed strong antibunching correlations (data not shown), an unequivocal signature of the single-molecule nature of the peaks shown in Fig. 1c.

To use the non-classical fluorescence properties of the single molecule directly, we trigger the single-photon emission by pumping with a periodic, pulsed laser source (pulse width $\tau_p = 35\ \text{ps}$, repetition rate $\nu = 6.25\ \text{MHz}$). Single photons are then generated at predetermined times, within the accuracy of the emission lifetime of a few nanoseconds. We used the standard time-correlated single-photon counting technique to measure the average time structure of our source (see Fig. 2). The inset shows that the photons are emitted at times separated by the laser repetition time (160 ns). As expected, the main figure shows that the excited state population rises in a very short time, demonstrating the rapid pumping of the molecule to its emitting state. The distribution then decays exponentially with a time constant of $\tau_f = 3.8\ \text{ns}$, consistent with the lifetime of terrylene in p-terphenyl as deduced from the homogeneous width of approximately 40 MHz at low temperature²³.

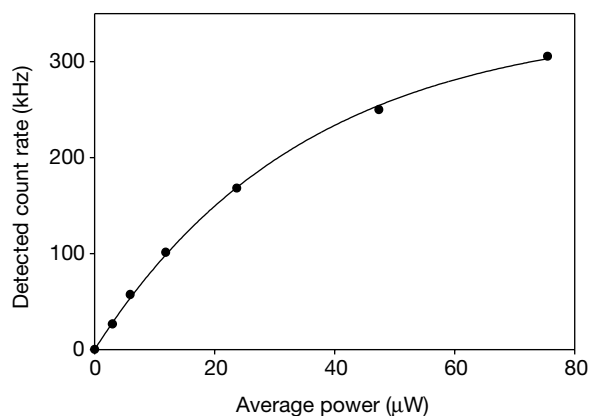


Figure 3 Determination of $p(1)$. The figure shows the intensity dependence of the emission rate from the triggered single photon source (dots) with a two-level model fit to the saturation behaviour (solid line). The maximum measured count rate is 310 kHz, and the expected saturation count rate from the fit is $S_\infty = 343\ \text{kHz}$, with $I_s = 1.2\ \text{MW cm}^{-2}$.

We now discuss the probabilities $p(m)$ that m photons will be emitted by the molecule after each pulse. Since the pump pulse width is very short compared to the fluorescence lifetime, the probability that two photons will be emitted per pulse, $p(2)$, is very small (see below). Thus the detected count rate from the molecule, S , is directly proportional to the probability $p(1)$ that a single photon is emitted after each laser pulse: $S = \eta \nu p(1)$ where η is the detection efficiency. By consideration of the losses in filters and optics as well as the collection efficiency for a molecular dipole oriented nearly perpendicular to the plane of the microscope stage²⁴, we find $\eta \approx 6\%$. To estimate the maximum value of $p(1)$, we performed a power saturation study of the detected count rate. In Fig. 3 we show the measured S as a function of the average excitation power of the pulsed laser. The signal shows power saturation behaviour that is well fitted by the saturation law of a two-level system from rate equations, $S = S_\infty (I/I_s) / (1 + I/I_s) [1 - \exp\{-(1 + I/I_s)\tau_p/\tau_f\}] = S_\infty p(1)$ where I and I_s are respectively the excitation and the saturation intensities. From the fit we obtain a saturation count rate $S_\infty = 343 \pm 6\ \text{kHz}$ and $I_s = 1.2 \pm 0.1\ \text{MW cm}^{-2}$ (in agreement with continuous-wave measurements, not shown). The maximum measured count rate in our experiment reached 310 kHz and was limited only by the available laser power. The maximum achieved $p(1)$ can be determined in two ways: using our maximum pumping intensity and I_s determined above, $p(1) = 88\%$; using the maximum observed S and η , $p(1) = 83\%$, in good agreement.

A small value of double-photon emission, $p(2)$, is essential to the usefulness of a single-photon source. For example, quantum key distribution using polarization states of photons requires immunity from a beam-splitter attack²⁵, where a large value of $p(2)$ allows an intruder to obtain information without being detected. Two-photon emission requires that the molecule must first be excited, then a photon must be emitted, and then the molecule must be

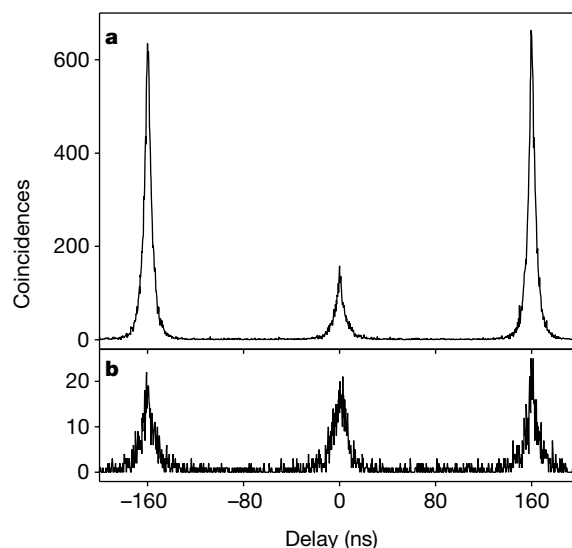


Figure 4 Second-order intensity correlation function of the emitted photons. The traces show the histogram of delays between detected photons under pulsed excitation with $6\ \mu\text{W}$ average power measured with the start–stop apparatus of Fig. 1d. **a**, A single molecule is excited and the ratio of the area of the central peak to the lateral ones is 0.27. The area of the central peak is proportional to $(B^2 + 2SB)$, where B is the average rate detected from the background. In addition to the background–background events, a larger contribution arises from the signal–background and background–signal events for the start–stop coincidences. Similarly, the area of each lateral peak is proportional to $(B + S^2)$. The ratio of the area of the central peak to the lateral peak is in good quantitative agreement with the ratio expected from the measured average count rate for the signal and the background ($S/B \approx 6$). **b**, When the laser spot is positioned away from any molecule, all peaks have the same area, as expected for a Poissonian source. The data were accumulated for **a**, 120 s and **b**, 300 s.

excited a second time, all within 35 ps. We have estimated the value of $p(2)$ and find it to be less than 8×10^{-4} at our maximum intensity.

To directly demonstrate the non-classical sub-poissonian statistics of our source, we measured the intensity correlation function of the emitted light with the coincidence set-up of Fig. 1d. For both the single-molecule (Fig. 4a) and background (Fig. 4b) cases, the histograms show the expected peak pattern, given the time pattern of the photon emission in the inset of Fig. 2. For poissonian light, such as that from an attenuated pulsed laser or the fluorescence from the background excited by the laser pulses, the central peak is identical in intensity and shape to the lateral ones (Fig. 4b). In the case of a single molecule, the central peak should vanish altogether as no more than one single photon can be emitted by the molecule. The ratio of the central peak's area to the area of the lateral peaks is the signature of the sub-poissonian statistics of the light emitted by our source. The residual peak at zero delay in Fig. 4a arises from coincidence events involving background photons excited during each laser pulse. In our experiment, the background signal shows a lifetime of about 4 ns, which indicates that it arises from weak fluorescence from out-of-focus terrylene molecules, not from Raman scattering.

It is convenient to compare the probability distribution $p(m)$ for our source to that expected from a Poisson distribution by means of the Mandel parameter $Q_s = (\sigma^2 - n_{av})/n_{av}$, where σ^2 is the variance of the distribution and n_{av} is the average number of photons²⁶. At the highest pumping power, the probabilities of our source are $p(0) = 0.14$, $p(1) = 0.86$ and $p(m > 1) \approx 0$. This distribution is radically different from that for a pulsed coherent source with the same $n_{av} = 0.86$: $p_{coh}(0) = 0.42$, $p_{coh}(1) = 0.36$, $p_{coh}(2) = 0.16$, The Mandel parameter of our source is $Q_s = -0.86$, not far from -1, the value expected for a perfect single-photon emitter, and far from 0, the value for a poissonian source. The Mandel parameter Q_d of the detected photon counts is naturally affected by the light detection efficiency²⁷. Using $Q_d = Q_s/(\eta/2)$, we find $Q_d \approx -3\%$.

The parameters of our source (repetition rate and single-photon generation probability) are limited only by the laser system used; nevertheless, the current performance already surpasses that of previous work. This high performance combined with the simplicity of our source may make it suitable for a variety of quantum optical experiments and for other applications where triggered single photons are needed. Photons are emitted into a range of solid angles, which can limit the detection efficiency; however, optical solutions to this problem can be envisaged, that is, one can imagine that the single molecule could be coupled to a single cavity mode to reduce losses, change the emission pattern, or modify the emission lifetime and thus increase the emission rate. To reduce the background (from out-of-focus molecules or Raman scattering), reduced terrylene doping, pumping with z-axis polarized light, or use of a crystalline system with a more favourable orientation of the single absorber can be utilized. With further development, single molecules in solids may soon provide compact and reliable sources of single photons. □

Received 27 January; accepted 15 August 2000.

1. Meystre, P. & Sargent, M. III *Elements of Quantum Optics* (Springer, Berlin, 1991).
2. Drummond, P. D. *Quantum Squeezing* (Springer, Berlin, 1999).
3. Special issue on quantum information. *Phys. World* 11(3), (1998).
4. Bennett, C. H., Brassard, G. & Ekert, A. K. Quantum cryptography. *Sci. Am.* 267(4), 50–57 (1992).
5. Turchette, Q. A., Hood, C. J., Lange, W., Mabuchi, H. & Kimble, H. J. Measurement of conditional phase shifts for quantum logic. *Phys. Rev. Lett.* 75, 4710–4713 (1995).
6. Parkins, A. S., Marte, P., Zoller, P., Carnal, O. & Kimble, H. J. Quantum-state mapping between multilevel atoms and cavity light fields. *Phys. Rev. A* 51, 1578–1596 (1995).
7. Cirac, J. I., Zoller, P., Kimble, H. J. & Mabuchi, H. Quantum state transfer and entanglement distribution among distant nodes in a quantum network. *Phys. Rev. Lett.* 78, 3221–3224 (1997).
8. Kuhn, A., Hennrich, M., Bando, T. & Rempe, G. Controlled generation of single photons from a strongly coupled atom-cavity system. *Appl. Phys. B* 69, 373–377 (1999).
9. Imamoglu, A., Schmidt, H., Woods, G. & Deutsch, M. Strongly interacting photons in a nonlinear cavity. *Phys. Rev. Lett.* 79, 1467–1470 (1997).

10. Imamoglu, A. & Yamamoto, Y. Turnstile device for heralded single photons: coulomb blockade of electron and hole tunneling in quantum confined p-i-n heterojunctions. *Phys. Rev. Lett.* 72, 210–213 (1994).
11. Kim, J., Benson, O., Kan, H. & Yamamoto, Y. A Single-photon turnstile device. *Nature* 397, 500–503 (1999).
12. Brunel, C., Lounis, B., Tamarat, P. & Orrit, M. Triggered source of single photons based on controlled single molecule fluorescence. *Phys. Rev. Lett.* 83, 2722–2725 (1999).
13. Xie, X. S. & Trautman, J. K. Optical studies of single molecules at room temperature. *Annu. Rev. Phys. Chem.* 49, 441–480 (1998).
14. Moerner, W. E. & Orrit, M. Illuminating single molecules in condensed matter. *Science* 283, 1670–1676 (1999).
15. Ambrose, W. P., Goodwin, P. M., Martin, J. C. & Keller, R. A. Single-molecule detection and photochemistry on a surface using near-field optical-excitation. *Phys. Rev. Lett.* 72, 160–163 (1994).
16. Ambrose, W. P. *et al.* Fluorescence photon antibunching from single molecules on a surface. *Chem. Phys. Lett.* 269, 365–370 (1997).
17. De Martini, F., Di Giuseppe, G. & Marrocco, M. *Phys. Rev. Lett.* 76, 900–903 (1996).
18. Kistner, S. C., Jonsson, P., Rarity, J. G. & Tapster, P. R. Intensity fluctuation spectroscopy of small numbers of dye molecules in a microcavity. *Phys. Rev. A* 58, 620–627 (1998).
19. Fleury, L., Sick, B., Zumofen, G., Hecht, B. & Wild, U. P. High photo-stability of single molecules in an organic crystal at room temperature observed by scanning confocal optical microscopy. *Mol. Phys.* 95, 1333–1338 (1998).
20. Kulzer, F., Koberling, F., Christ, T., Mews, A. & Basche, T. Terrylene in p-terphenyl: single molecule experiments at room temperature. *Chem. Phys.* 247, 23–34 (1999).
21. Basché, T., Moerner, W. E., Orrit, M. & Talon, H. Photon antibunching in the fluorescence of a single dye molecule trapped in a solid. *Phys. Rev. Lett.* 69, 1516–1519 (1992).
22. Fleury, L., Segura, J.-M., Zumofen, G., Hecht, B. & Wild, U. P. Nonclassical photon statistics in single molecule fluorescence at room temperature. *Phys. Rev. Lett.* 84, 1148–1151 (2000).
23. Kummer, S., Basché, T. & Bräuchle, C. Terrylene in p-terphenyl: a novel single crystalline system for single molecule spectroscopy at low temperatures. *Chem. Phys. Lett.* 229, 309–316 (1994).
24. Plakhotnik, T., Moerner, W. E., Palm, V. & Wild, U. P. Single molecule spectroscopy: maximum emission rate and saturation intensity. *Opt. Commun.* 114, 83–88 (1995).
25. Buttler, W. T. *et al.* Quantum key distribution over 1 km. *Phys. Rev. Lett.* 81, 3283–3286 (1998).
26. Short, H. & Mandel, L. Observation of sub-poissonian photon statistics. *Phys. Rev. Lett.* 51, 384–387 (1983).
27. Mandel, L. Sub-poissonian photon statistics in resonance fluorescence. *Opt. Lett.* 4, 205–207 (1979).
28. Miller, G. D. *et al.* 42%-efficient single-pass cw second-harmonic generation in periodically poled lithium niobate. *Opt. Lett.* 22, 1834–1836 (1997).

Acknowledgements

We thank Lightwave Electronics for use of the mode-locked Nd-YAG laser, M. Fejer and M. Fayer for providing non-linear crystals, and the Stanford Synchrotron Radiation Laboratory for providing the pulse height analyser. B. L. thanks D. Wright for assistance and NATO for Fellowship support.

Correspondence and requests for materials should be addressed to W.E.M. (e-mail: moerner@stanford.edu).

Logical computation using algorithmic self-assembly of DNA triple-crossover molecules

Chengde Mao*, Thomas H. LaBean†, John H. Reif† & Nadrian C. Seeman*

* Department of Chemistry, New York University, New York, 10003, USA &

† Department of Computer Science, Duke University, Durham, North Carolina 27707, USA

Recent work^{1–3} has demonstrated the self-assembly of designed periodic two-dimensional arrays composed of DNA tiles, in which the intermolecular contacts are directed by 'sticky' ends. In a mathematical context, aperiodic mosaics may be formed by the self-assembly of 'Wang' tiles⁴, a process that emulates the operation of a Turing machine. Macroscopic self-assembly has been used to perform computations⁵; there is also a logical equivalence between DNA sticky ends and Wang tile edges^{6,7}. This suggests that the self-assembly of DNA-based tiles could be used to perform DNA-based computation⁸. Algorithmic aperiodic self-assembly requires greater fidelity than periodic self-assembly,

because correct tiles must compete with partially correct tiles. Here we report a one-dimensional algorithmic self-assembly of DNA triple-crossover molecules⁹ that can be used to execute four steps of a logical (cumulative XOR) operation on a string of binary bits.

A variety of different DNA tile types have been used in previous assemblies, including double-crossover molecules¹, triple-crossover molecules⁹, and parallelograms produced from Holliday junction analogues³. Here we have used triple-crossover molecules; their

molecular structure is illustrated in Fig. 1a. The molecule contains four strands (shown in red, green, blue and purple) that self-assemble through Watson–Crick base pairing to produce three double helices in a roughly planar arrangement; each double helix is connected to adjacent double helical domains at two points where their strands cross over between them. The ends of the central double helix are closed by hairpin loops, but the other helices can terminate in sticky ends containing information that directs the assembly of the tiles.

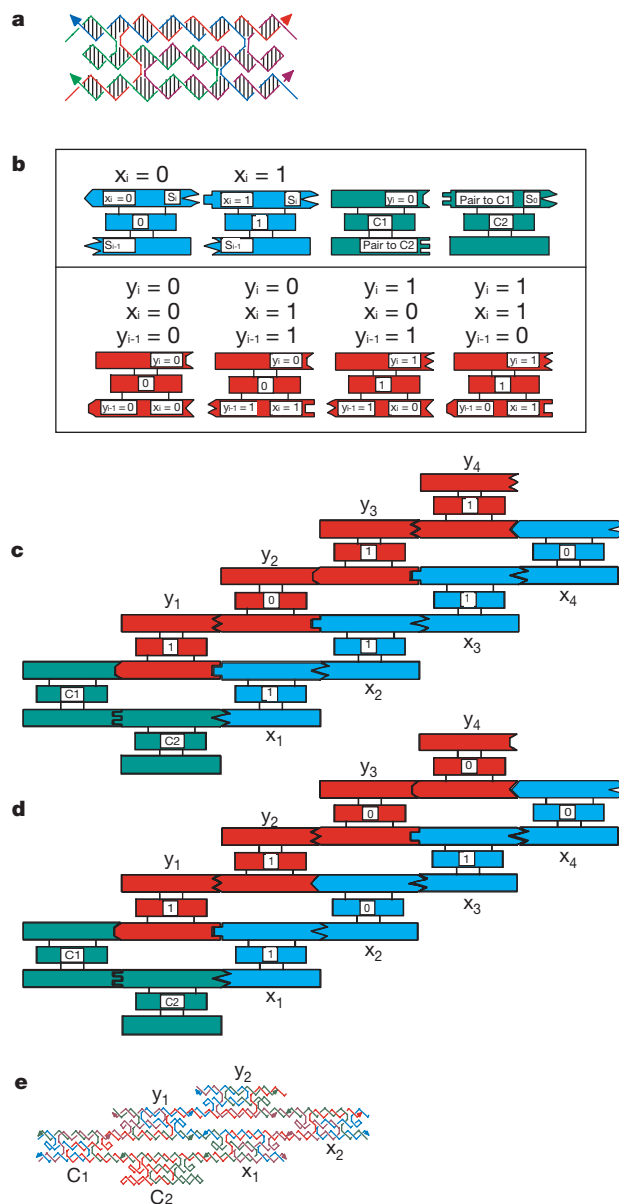


Figure 1 Calculation of cumulative XOR by self-assembly of DNA tiles. **a**, An individual triple-crossover tile. The four coloured backbone strands assemble to form three double-helical domains; arrowheads indicate 3' ends. **b**, Component tiles. The three helical domains are drawn as rectangles, flanked by sticky ends shown as geometrical shapes. The value of each tile is in the central rectangle. The meaning of each sticky end is also indicated. Shown are the two x tiles (blue), the four y tiles (red), and the initialization corner tiles, C1 and C2 (green). The y_i tiles are upside down from the x_i tiles. **c**, **d**, The two calculations performed. The values of each tile are the same as in **b**, and the sticky ends are the same, although the labels have been omitted for clarity. The tiles are shown assembled to perform the cumulative XOR calculation; note the complementarity of sticky-ended association at each molecular interface. All sticky ends are asymmetric, so that when the same meanings touch (for example, $x_i = 0$ pairing with $x_i = 0$), the sequences

are complementary, but not self-complementary. The operations are designed to proceed from lower left to upper right, because the x_i and C1 and C2 tiles have longer sticky ends than the y_i tiles. After the first y tile has been added, $x_1 = 1$ and $y_1 = 1$, for both arrays, because $y_1 = x_1$. In the next step, the array in **c** contains $x_2 = 1$: because $x_2 = y_1$, y_2 should be 0, and only the y tile with value 0 and inputs y_{i-1} and $x_i = 1$ fits properly between y_1 and x_2 . In **d**, $x_2 = 0$, so $x_2 \neq y_1$, and y_2 should be 1, as shown. The sequences are available in the Supplementary Information. **e**, The reporter strand. At the end of the assembly, the reporter strand running through the x diagonal array, around the corner, and then back up the y diagonal array is ligated, thereby associating the calculated output with the input. The strand structure in the vicinity of the corner is shown, and the reporter strand is drawn with a thick red line.

In periodic assemblies, the sticky ends contain positional information that directs the associations of one or more tile types to produce a periodic lattice. Here, the sticky ends serve the same role, but the self-assembly of the tiles is used to perform a computation, and the arrangement of the tiles does not display simple periodicity. Figure 1b shows the tiles used to perform the cumulative XOR calculation. The tiles are represented schematically; the three helices are depicted as connected rectangular forms terminating in sticky ends, which are represented as geometrical shapes, or non-cohesive blunt ends and hairpins, which are drawn flush.

The result of the XOR operation is a 0 if two input numbers are the same (two zeros or two ones), but it is 1 if one of the two numbers is 0 and the other is 1. The cumulative XOR consists of a series of Boolean inputs $x_1, x_2, x_3, \dots, x_n$, and the output is also a series of Booleans, $y_1, y_2, y_3, \dots, y_n$, where $y_1 = x_1$, and for $i > 1$, $y_i = y_{i-1} \text{ XOR } x_i$. The value of any y_i in these calculations also reports the even or odd parity of the first i values of x . Thus, two different kinds of input x tiles are needed, one whose value is 0 and a second whose value is 1. Chemically, the value of a tile, 0 or 1, is denoted by the presence of a restriction site: Pvu II (CAGCTG)

represents 0 and EcoR V (GATATC) represents 1.

The x tiles are shown in blue in Fig. 1b: Their value (0 or 1) is shown in their central rectangle, the upper-left sticky end reports this value, and the upper-right and lower-left sticky ends provide the means of connecting successive x tiles. These sticky ends are shown as geometrically complementary, as they would be for a general set of 16 parallel calculations. However, to demonstrate the efficacy of the procedure, the calculations performed here are definite four-bit calculations for which the order of the input x tiles was specified exactly by a series of different sticky ends.

Like x tiles, two values of y tiles are necessary, again representing 0 and 1. However, there are two ways to get each of these results: the value of a y tile can be 0 either because both inputs are 0 or because both are 1; likewise, the value of a y tile can be 1 because the value of one input is 0 and the other is 1, or vice versa. Thus, four different y tiles are needed.

The y tiles are shown in red in Fig. 1b. The tile values again are displayed on the central domain, and this value is reported by the sticky end on the right of the upper domain. The two inputs derive from the sticky ends on the left (y_{i-1}) and right (x_i) of the bottom domain. We note that the same sticky end in the input domain represents a given input, independent of the other end. For example, the right-side sticky end $x_i = 1$ has the same shape (sticky end) regardless of whether the left-side sticky end represents $y_{i-1} = 0$ (leading to a tile value of 1) or $y_{i-1} = 1$ (leading to a tile value of 0). There are only two different left sticky ends in the input (bottom) domain, and likewise only two different right sticky ends. Consequently, both sticky ends on each tile must pair correctly for the proper y_i tile to be inserted in the assembly. In contrast to periodic assembly, where correct tiles compete with incorrect tiles for each site in the lattice, here correct tiles are competing with partially correct tiles.

We have performed two different XOR-related self-assemblies, illustrated in Fig. 1c and d. In addition to the x and y tiles, two corner tiles, C1 and C2 (green in Fig. 1b) are used to initialize the two values of x_1 and y_1 , and to connect the input to the output. The self-assembly in Fig. 1c has the inputs $x_1 = x_2 = x_3 = 1$ and $x_4 = 0$. These correspond to output values of $y_1 = 1, y_2 = 0$, and $y_3 = y_4 = 1$. In a second self-assembly (Fig. 1d), $x_1 = 1, x_2 = 0, x_3 = 1$ and $x_4 = 0$, corresponding to $y_1 = y_2 = 1$ and $y_3 = y_4 = 0$. Note that for $i > 1$, the sticky ends on the bottom domain of each y_i tile complement those on the y_{i-1} tile on its left and the x_i tile on its right.

Once the self-assembly has occurred, it is necessary to extract the answer. For this purpose, each molecular tile contains a 'reporter strand'¹⁰, which traverses the tile in a diagonal pathway⁹; the reporter strand is illustrated as a thick red strand in the tile shown in Fig. 1a, which is an x -type tile. Following self-assembly, the reporter strands are ligated to each other to produce a long reporter strand that contains the inputs and outputs of the calculation. The ligated long reporter strand in the vicinity of the corner of the assembly is shown as a thick red strand on the molecular diagram in Fig. 1e.

The sticky ends used in the assembly of the C_1 - C_2 - x_1 - x_2 - x_3 - x_4 unit contain seven nucleotides, and the sticky ends used to include the y_i tiles in the assembly contain five nucleotides. The tiles were first assembled individually from their component strands by cooling slowly from 90 °C to room temperature, as done previously¹⁻³. 20- μ l aliquots of stock solutions (in USB ligation buffer) of both C tiles (100 nM), the four x tiles (100 nM), and the four y tiles (400 nM) were then combined and incubated for 30 min each at temperatures of 37, then 22, and finally 4 °C. During incubation and subsequent steps, 20- μ l aliquots of three double helices, each with a sticky end (one helix to pair with the free sticky end on x_4 and two helices to pair with the two possible free sticky ends on y_4), that contained radioactively labelled PCR primers (800 nM) were also present in solution.

Ligation was initiated by adding 20 units of T4 DNA ligase and,

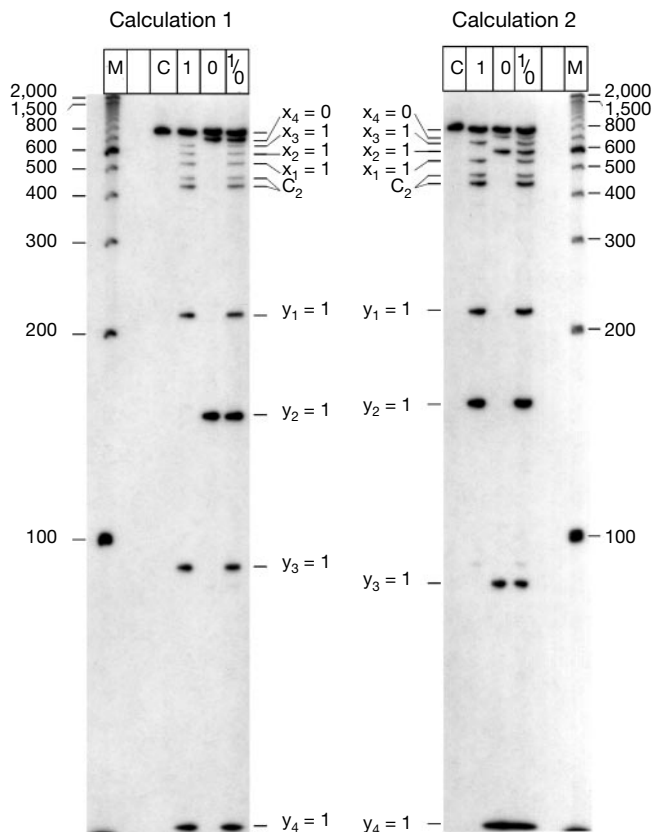


Figure 2 Denaturing gels illustrating cumulative XOR calculations. Two 6% denaturing gels are shown, corresponding to the two calculations described. M, lanes containing 100-mer marker ladders; C, lanes containing undigested reporter strands. The lanes digested with Pvu II are labelled 0, to indicate that sensitivity to this enzyme represents a tile with the value zero, and those digested with EcoR V are labelled 1, because the site for that enzyme represents a value of one. A lane combining the two individual digestions is labelled 1/0. The bands corresponding to the input values x_4, x_3, x_2 and x_1 are visible at the tops of the lanes, in descending order. The C2 tile contains a hairpin with an EcoR V site, so the site is present twice in the reporter strand at that location. The bands corresponding to the output values y_1, y_2, y_3 and y_4 are visible at the bottom of the gel. Faint bands corresponding to erroneous Pvu II cleavage are visible near y_2 in calculation 1 and y_3 in calculation 2. A faint band corresponding to erroneous EcoR V cleavage is present near y_3 in calculation 1.

over a 3-hour period, the solution was brought slowly to 16 °C as ligation proceeded. The strand was amplified by polymerase chain reaction (PCR), using the primers that were ligated to each end of the long reporter strand. A strand of the proper length was eluted from a denaturing gel, was re-annealed, and was subjected to restriction by either of the restriction enzymes. The results are displayed in Fig. 2. The answer produces a barcode display, much like that used in ref. 11 to visualize the answers to RNA computation of chess problems. The correct answers are evident as dark bands in the gels shown in Fig. 2.

Small proportions of incorrect bands are visible on the original gels for both calculations. Thus, in lane 1 (EcoR V) of calculation 1, there is an incorrect band at the 0 position of y_2 , and a similar incorrect band in the 0 lane at y_3 is also present. However, if the array fills from the corner, some of the incorrect y_3 intensity could result from 'correct' molecules propagating the previous error. Similarly, calculation 2 has one detectable error band resulting from cleavage of EcoR V at y_3 .

We estimate the error level to be about 2–5%, but quantitative error analysis is complicated by differential cleavage activities between the two enzymes, combined with the possibility of star-activity (sequence infidelity) and probable multiple cleavage of the same strand. Also, the individual enzymes cleave with different activities at different sites, as seen in the differential cleavage of the two EcoR V sites in the C2 tile. We may have reduced our observation of self-assembly errors by selecting only those tiles that ligated correctly, because the enzyme specificity for exact pairing, although imperfect¹², may have performed some discrimination for the system. A previous two-molecule, single-step competition experiment estimated error rates below 1.6% (ref. 13).

The algorithmic molecular assembly described here demonstrates a non-trivial DNA computation done by self-assembly. Examples of SAT (satisfaction) problems solved in a DNA context^{11,14,15} entailed laboratory operations for each clause in a logical statement, whereas a single self-assembly step is used here. This suggests that computation by self-assembly may be scalable. Another recent work¹⁶ also uses only a single assembly step, but its scalability relies on proper hairpin formation in very long single-stranded molecules.

XOR computation on pairs of bits (as done here) can be used for executing a one-time pad cryptosystem that provides theoretically unbreakable security¹⁷. Other applications could involve the algorithmically directed self-assembly of intricate patterns and smart materials. We used y tiles repetitively in both assemblies, and would need no more species of y tiles, regardless of the length of the calculation. Thus, if the assembly principles applied here can be extended to two and three dimensions, it will be possible to prepare nanoscale patterns and smart materials by laying out components algorithmically, without the need to specify and prepare a unique element for every position of the array.

By using more nucleotides in the sticky ends of the input tiles than the output tiles, we have used the principle of 'frames'^{6,7,18}. This feature performs the computation in the presence of a well-defined border. Such borders are likely to be useful, because they set limits on the extent of the calculation or patterning; combining framed arrays will facilitate a modular approach to the process. □

Received 27 April; accepted 3 August 2000.

1. Winfree, E., Liu, F., Wenzler, L. A. & Seeman, N. C. Design and self-assembly of two-dimensional DNA crystals. *Nature* **394**, 539–544 (1998).
2. Liu, F., Sha, R. & Seeman, N. C. Modifying the surface features of two-dimensional DNA crystals. *J. Am. Chem. Soc.* **121**, 917–922 (1999).
3. Mao, C., Sun, W. & Seeman, N. C. Designed two-dimensional DNA Holliday junction arrays visualized by atomic force microscopy. *J. Am. Chem. Soc.* **121**, 5437–5443 (1999).
4. Wang, H. in *Proceedings of a Symposium in the Mathematical Theory of Automata* 23–26 (Polytechnic Press, New York, 1963).
5. Rothmund, P. W. K. Using lateral capillary forces to compute by self-assembly. *Proc. Natl. Acad. Sci. USA* **97**, 984–989 (2000).
6. Winfree, E. in *DNA Based Computers: Proceedings of a DIMACS Workshop, April 4, 1995, Princeton University* (eds Lipton, R. J. & Baum, E. B.) 199–221 (American Mathematical Society, Providence, RI, 1996).

7. Winfree, E. *Algorithmic Self-Assembly of DNA*. PhD Thesis, Caltech (1998).
8. Adleman, L. Molecular computation of solutions to combinatorial problems. *Science* **266**, 1021–1024 (1994).
9. LaBean, T. *et al.* The construction, analysis, ligation and self-assembly of DNA triple crossover complexes. *J. Am. Chem. Soc.* **122**, 1848–1860 (2000).
10. Seeman, N. C. Nucleic acid nanostructures and topology. *Angew. Chem. Int. Edn Engl.* **37**, 3220–3238 (1998).
11. Faulhammer, D., Cukras, A. R., Lipton, R. J. & Landweber, L. F. Molecular computation: RNA solutions to chess problems. *Proc. Natl. Acad. Sci. USA* **97**, 1385–1389 (2000).
12. Harada, K. & Orgel, L. E. Unexpected substrate specificity of T4 DNA ligase revealed by *in vitro* selection. *Nucleic Acids Res.* **21**, 2287–2291 (1993).
13. Winfree, E., Yang, X. & Seeman, N. C. in *DNA Based Computers: II Proceedings of a DIMACS Workshop, June 10–12, 1996, Princeton University* (eds Landweber, L. F. & Baum, E. B.) 217–254 (American Mathematical Society, Providence, RI, 1999).
14. Liu, Q. *et al.* DNA computing on surfaces. *Nature* **403**, 175–179 (2000).
15. Pirrung, M. C. *et al.* The arrayed primer extension method for DNA microchip analysis. Molecular computation of satisfaction problems. *J. Am. Chem. Soc.* **122**, 1873–1882 (2000).
16. Sakamoto, K. *et al.* Molecular computation by DNA hairpin formation. *Science* **288**, 1223–1226 (2000).
17. Gehani, A., LaBean, T. H. & Reif, J. H. in *DNA Based Computers: Proceedings of a DIMACS Workshop, June 1999, MIT* (ed. E. Winfree) (DIMACS Series in Discrete Mathematics and Theoretical Computer Science, American Mathematical Society, Providence, RI, in the press).
18. Reif, J. H. in *DNA Based Computers: III Proceedings of a DIMACS Workshop, June 23–25, 1997, University of Pennsylvania* (eds Rubin, H. & Wood D. H.) 217–254 (American Mathematical Society, Providence, RI, 1999).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank E. Winfree and A. Carbone for valuable discussions. This work has been supported by grants from DARPA and the National Science Foundation to J.H.R. and N.C.S.; ONR, USAF, NSF and NIH grants to N.C.S.; and NSF and ARO grants to J.H.R.

Correspondence and requests for materials should be addressed to N.C.S. (e-mail: ned.seeman@nyu.edu).

Nano-sized transition-metal oxides as negative-electrode materials for lithium-ion batteries

P. Poizot, S. Laruelle, S. Grugeon, L. Dupont & J.-M. Tarascon

Laboratoire de Réactivité et Chimie des Solides, Université de Picardie Jules Verne, CNRS UPRES A 6007, 33 rue Saint Leu, F-80039, Amiens, France

Rechargeable solid-state batteries have long been considered an attractive power source for a wide variety of applications, and in particular, lithium-ion batteries are emerging as the technology of choice for portable electronics. One of the main challenges in the design of these batteries is to ensure that the electrodes maintain their integrity over many discharge–recharge cycles. Although promising electrode systems have recently been proposed^{1–7}, their lifespans are limited by Li-alloying agglomeration⁸ or the growth of passivation layers⁹, which prevent the fully reversible insertion of Li ions into the negative electrodes. Here we report that electrodes made of nanoparticles of transition-metal oxides (MO, where M is Co, Ni, Cu or Fe) demonstrate electrochemical capacities of 700 mA h g^{−1}, with 100% capacity retention for up to 100 cycles and high recharging rates. The mechanism of Li reactivity differs from the classical Li insertion/deinsertion or Li-alloying processes, and involves the formation and decomposition of Li₂O, accompanying the reduction and oxidation of metal nanoparticles (in the range 1–5 nanometres) respectively. We expect that the use of transition-metal nanoparticles to enhance surface electrochemical reactivity will lead to further improvements in the performance of lithium-ion batteries.

Swagelok-type cells¹⁰ were assembled and cycled using a Mac-Pile automatic cycling/data recording system (Biologic Co, Claix, France) between 3 and 0.01 V. These cells comprise (1) a 1-cm², 75-μm-thick disk of composite positive electrode containing 7–10 mg of MO (from Aldrich or Union Minière, unless otherwise specified) mixed with 10% of carbon SP (Black carbon from MM, Belgium), and made according to Bellcore's plastic Li-ion technology¹¹, (2) a 1-cm² Li metal disk as the negative electrode, and (3) a Whatman GF/D borosilicate glass fibre sheet saturated with a 1 M LiPF₆ electrolyte solution in 1:1 dimethyl carbonate:ethylene carbonate as separator/electrolyte. The cells were cycled at a rate of C/5 (that is, one lithium per formula unit in 5 hours).

The voltage–composition traces of the MO/Li cells made with Co, Ni or Fe oxide (Fig. 1a) show some similarities. During the first discharge, the potential rapidly drops to reach a plateau (the potential and the amplitude depend on M), and then continuously decreases down to 0.01 V. The amplitudes of both the plateau and the sloping part of the curve are, for example, about 2 and 0.7 Li per M, respectively, for CoO. On the following charge, about 2 Li per M could be removed, leading to reversible capacities ranging from 600 to 800 mA h per g of MO; these values are about twice those of today's graphite negative electrodes. The second discharge curve differs considerably from the first, suggesting drastic, lithium-driven, structural or textural modifications.

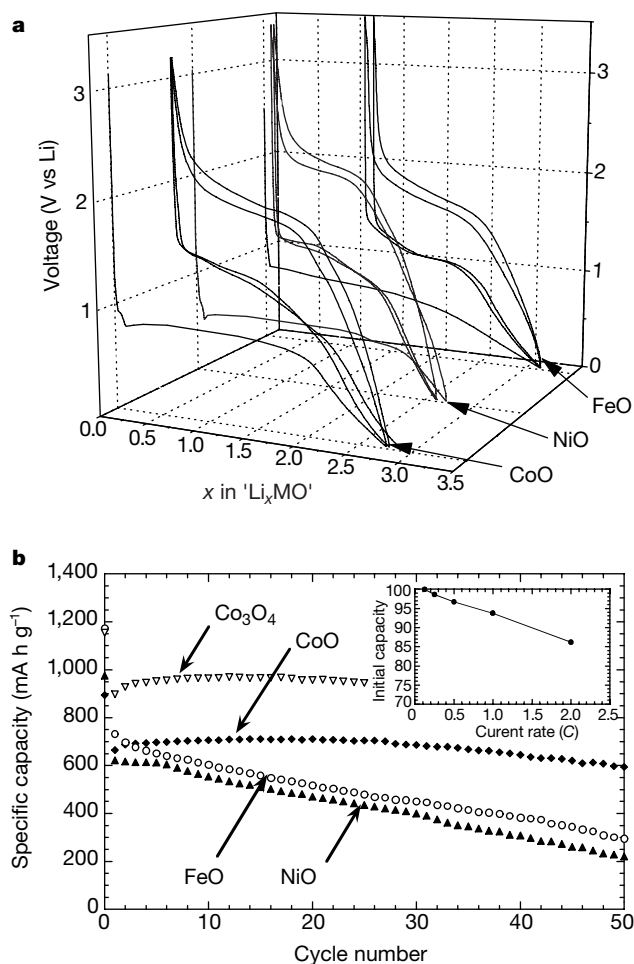


Figure 1 Properties of MO/Li cells. **a**, The voltage–composition profile for various MO/Li cells cycled between 0.01 V and 3 V at a rate of C/5 (1 lithium in 5 hours). The capacity fading for the same cells under similar conditions are shown in **b**, where we also show data for a Co₃O₄/Li cell in order to show that the reported behaviour is not specific to divalent oxides. Inset, the rate capability of a CoO electrode.

The capacity of the MO/Li cells on cycling showed two types of variation (Fig. 1b): for FeO and NiO, the reversible capacity continuously decayed, whereas for Co oxides the capacity remained constant or even slightly increased, finally decaying after *n* cycles (data not shown), *n* being dependent on the particle morphology, as will be discussed later. We found that the capacity retention was at its best when cycling was done at room temperature over the entire (3.0–0.01 V) voltage range. These metal oxide electrodes were found to sustain good rate capabilities, being able to deliver 85% of their total capacity at a 2C rate (see Fig. 1b inset).

The evolution of the crystallinity of the electrode material on cycling was determined by means of an *in situ* X-ray diffraction (XRD) electrochemical cell¹² (assembled similarly to our Swagelok cell but having a beryllium window as current collector on the MO side) that was placed on a SCINTAG diffractometer, and connected to the Mac-Pile system. Pure CoO adopts a structure of the rock-salt type, with *a* = 4.258 Å. Because the most intense lines are (111) and (200), with values of 2θ of 36.52° and 42.42°, respectively, we limited our 2θ data collection scan to the 35–45° range. As Li reacted with CoO (Fig. 2), we observed a continuous decrease in the CoO Bragg peaks that totally vanished when *x* (in the formula Li_{*x*}MO) equals 2, leading to a featureless XRD pattern that did not evolve upon further reduction down to 0.01 V or subsequent oxidation up to 3.0 V. The discharge plateau at 0.8 V obtained at a C/10 rate (Fig. 2 inset) can then be ascribed to a decomposition/amorphization of the CoO material. As exactly two electrons are consumed during this reduction plateau (as separately and accurately determined by potentiodynamic intermittent titration), we suggest that CoO reacts with Li according to the following reaction: CoO + 2 Li⁺ + 2e⁻ → Co + Li₂O. This leads to the formation of metallic Co and Li₂O nanoparticles that are presumably smaller than the X-ray coherence length, as there are no diffraction peaks.

To investigate more closely the lithium-driven structural and morphological changes, we studied CoO-based electrodes at various stages of the reduction and oxidation processes by means of a Philips CM12 transmission electron microscope (TEM). To perform such experiments, cells were cycled down to the required

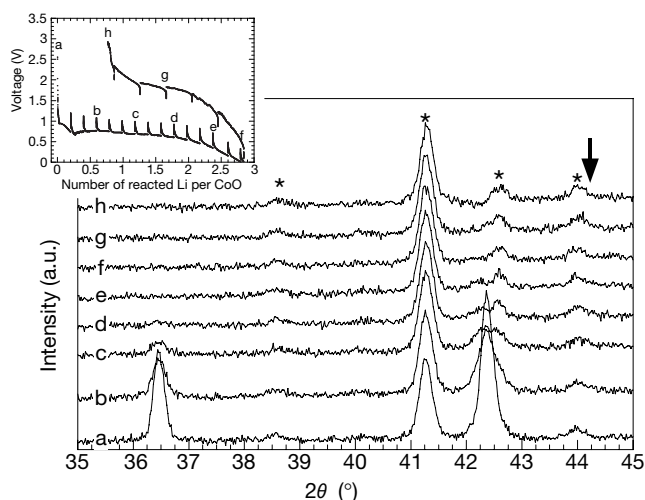


Figure 2 *In situ* X-ray diffraction patterns collected at various states of discharge and charge of a CoO/Li electrochemical cell. Inset, the voltage–composition profile for such a cell, where the letters a to h denote the *x* values (in Li_{*x*}MO) at which the corresponding X-ray patterns were taken. Such an experiment was performed in a galvanostatic intermittent mode, using current-on periods of two hours at a rate of C/10, separated by two-hour open-circuit periods during which X-ray data were taken. The peaks marked with an asterisk correspond to Be and BeO, and the arrow indicates where a Bragg peak corresponding to metallic Co should appear.

voltage and opened in a dry box; the partially lithiated material was recovered and washed with dimethyl carbonate (DMC) before being placed onto a copper grid mounted on our TEM sample holder. Through a special mobile airlock of our own design, the sample was then transferred to the TEM for examination without any exposure to air. We then examined the selected-area electron diffraction (SAED) pattern and bright-field images of the samples. For reasons of conciseness, we will only report here results for the initial material, and the totally discharged (reduced) and the fully recharged (oxidized) materials.

Bright-field images (Fig. 3a) of the initial CoO powders indicate that they consist of particles having dimensions ranging from 1,000 to 2,000 Å, while the SAED pattern (Fig. 3b) taken along the $[22-4]^*$ direction (the asterisk indicates that the direction is given in the reciprocal space) indicates well crystallized particles. When CoO is fully reduced by lithium, the bright-field image (Fig. 3c) shows, despite the fact that the overall shape of the starting particle is preserved, a complete disintegration of the starting 1,000 Å particles into 10–20 Å metallic nanoparticles. These Co nanoparticles are dispersed in a lithia (Li_2O) matrix, with the Li_2O + nanoparticles being surrounded by a solid electrolyte interface, the nature and composition of which varied on cycling. The presence of weak and large rings in the SAED pattern (Fig. 3d) for the fully lithiated sample is evidence for the loss of sample crystallinity, in agreement with the XRD data. The nano-sized and pseudo-amorphous characters of the composite electrode, once created during the first discharge, are preserved on the following charge—as shown by the bright-field image (Fig. 3e) and SAED pattern (Fig. 3f). In contrast, its composition changes from Co to CoO (as deduced from SAED), indicating a reversible process, which was further confirmed by magnetic measurements capable of distinguishing between

ferromagnetic Co and antiferromagnetic CoO , and by *in situ* extend X-ray absorption fine structure (EXAFS) measurements as a function of cycling (data not shown).

If the nano-size of the metal oxide particles is the reason for their reactivity towards lithium, the capacity retention of such electrode materials should be extremely sensitive to their degree of division during the formation cycle, and hence to the precursor morphology. To check this, we performed a preliminary study on the Cu_2O system. Two batches of monodisperse Cu_2O powders, having particle sizes of 1 μm and 0.15 μm , were prepared according to

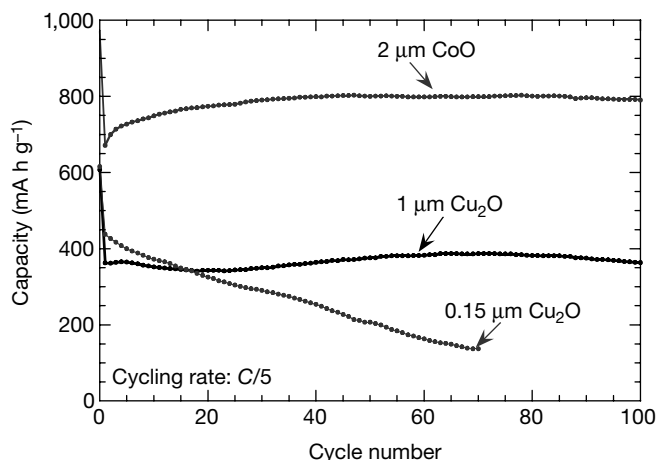


Figure 4 Capacity fading of Cu_2O -based electrodes as a function of the particle size. We also show data for an electrode based on 2- μm CoO powder, which showed 100% capacity retention.

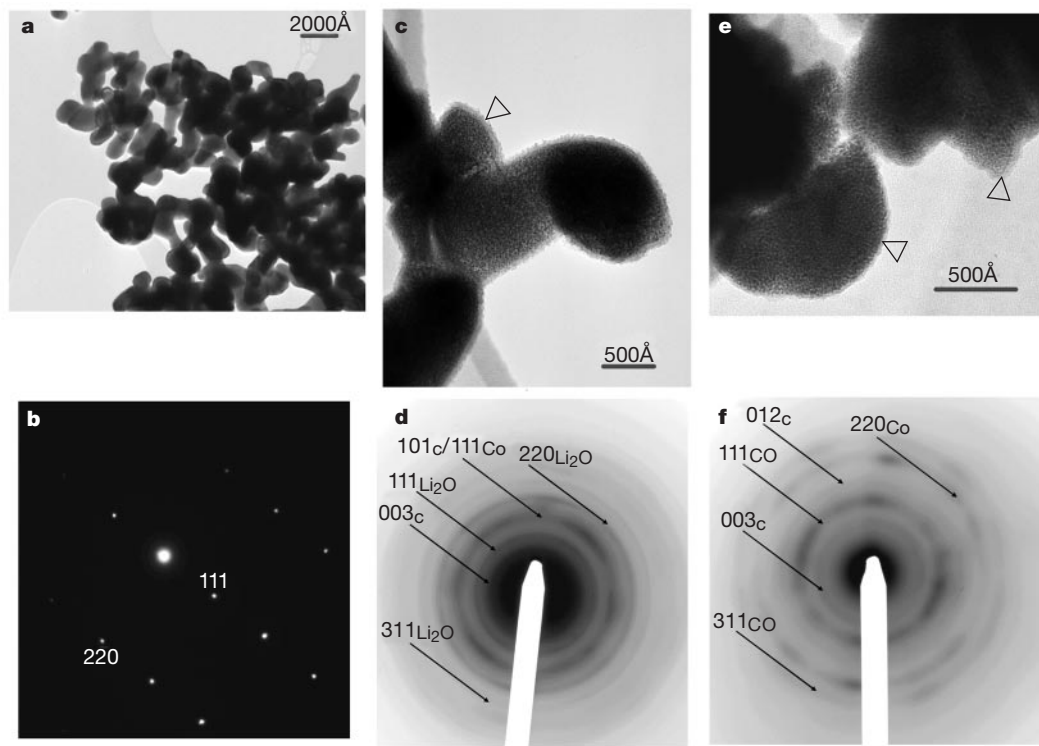


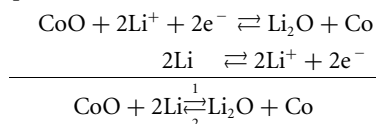
Figure 3 TEM images and SAED patterns of CoO electrodes taken from non-cycled, fully discharged and fully charged cells. **a**, Micrograph of the starting CoO electrode (1,000–2,000-Å particles are assembled in 1–2- μm aggregates). **b**, SAED patterns along $[22-4]^*$ showing the monolithic character of the CoO starting particles. **c**, Micrograph of the fully lithiated CoO electrode. Initial particles are transformed into 20-Å nanoparticles. The global shape is conserved; the white triangle indicates the “inorganic” solid electrolyte interface surrounding the agglomerates. **d**, The corresponding SAED pattern

showing the presence of Li_2O and Co inside the agglomerate. We note the pseudo-amorphous character as shown by the SAED pattern. The pattern is presented in its negative form in order to be able to easily observe the diffuse rings. **e**, Micrograph of a delithiated CoO electrode. The particle size is unaffected. **f**, The corresponding SAED pattern. Weak and diffuse rings, indexed using CoO, are observed. The subscript c in the hkl notation refers to carbon.

the polyol process¹³ (solutions of 2 g and 0.245 g of $\text{Cu}(\text{NO}_3)_2 \cdot 3\text{H}_2\text{O}$ in 75 ml of ethylene glycol were reduced under reflux at 160 °C for 3 h). Figure 4 shows how such a change in the particle size of the precursor can strongly affect the capacity retention; this suggests that for each metal oxide system there is an optimum precursor particle size, producing the best state of division of the metal particles and hence the best electrochemical performance. For example, 100% capacity retention after 100 cycles was achieved using 2- μm CoO particles that we had prepared.

Finally, we assembled $\text{LiMn}_2\text{O}_4/\text{CoO}$ cells. These cells were shown to have an average voltage of 2.2 V, and to deliver 95% of their total capacity from 3.0 to 1.2 V (the latter value is the expected cut-off voltage for future electronic devices) while exhibiting a good capacity retention at least over the first 40 cycles. The advantages of such MO-based lithium-ion cells over commercial LiCoO_2/C cells are due to the fact that MO has about twice the capacity of carbon per unit mass, and three times its density; MO therefore has about 6 times the capacity of carbon per unit volume.

The results we report here are of interest with respect to possible applications; they are also of interest at a fundamental level. The mechanism of the reaction with lithium of these metal oxides differs from the classical mechanisms, which are based either on reversible insertion/deinsertion of lithium into host structures or on lithium alloying reactions. This difference is due to the facts that most of these materials (CoO for instance) crystallize in a rock-salt structure that does not contain any available empty sites for Li ions, and also that none of the 3d metals considered forms alloys with Li. From our XRD and TEM observations, and the magnetic measurements, it appears that the reversible electrochemical reaction mechanism of Li with the transition-metal oxides, such as CoO , entails for the most part a displacive redox reaction, as follows.



This implies the reversible formation (reaction (1)) and decomposition (reaction (2)) of Li_2O .

Reaction (1) is thermodynamically feasible and thereby expected. In contrast, the ability to drive reaction (2) electrochemically is surprising. Indeed, Li_2O has always been reported to be electrochemically inactive, as further confirmed by our own failed attempts to electrochemically decompose Li_2O powders even when mechanically milled with Co powders. However, caution has to be exercised since, when dealing with nano-sized materials, chemical and physical phenomena are strongly affected. (For example, the melting point of gold can vary from its well known temperature of 1,064 °C to less than 200 °C when the particles become smaller than 20 Å (ref. 14).) On this basis, the electrochemically driven size confinement of the metal particles is thus believed to enhance their electrochemical activity towards the formation/decomposition of Li_2O . With decreasing particle size, an increasing proportion of the total number of atoms lies near or on the surface (for example, for particle sizes of 100 Å and 50 Å, this proportion is about 15–30%, and 30–60%, respectively) making the electrochemical reactivity of the particles more and more important. Such considerations can explain why the cycling performances of such materials should be extremely sensitive to their degree of division or aggregation.

In this work we have reported the reaction of nano-sized transition-metal oxides with Li^+ in the solid state. We are at present investigating the conditions needed to optimize the reactivity of such nanoparticles in non-aqueous Li-bearing electrolytes. □

Received 17 April; accepted 6 July 2000.

- Shodai, T., Okada, S., Tobishima, S. & Yamabi, J. Study of $\text{Li}_{1-x}\text{M}_x\text{N}$ (M = Co, Ni or Cu) system for use as anode in lithium rechargeable cells. *Solid State Ionics* **786**, 86–88 (1996).
- Takeda, Y. et al. Lithium secondary batteries using a lithium cobalt nitride, $\text{Li}_2\text{Co}_0.4\text{N}$, as the anode. *Solid State Ionics* **130**, 61–69 (2000).

- Idota, Y., Kubota, T., Matsufuji, A., MacKawa, Y. & Miyasaka, T. Tin-based amorphous oxide: A high-capacity lithium-ion storage material. *Science* **276**, 1395–1397 (1997).
- Kepler, K. D., Vaughey, J. T. & Thackeray, M. M. $\text{Li}_2\text{Cu}_2\text{Sn}_5$ ($0 < x < 13$): An intermetallic insertion electrode for rechargeable lithium batteries. *Electrochem. Solid State Lett.* **7**, 307–309 (1999).
- Mao, O., Dunlap, R. A. & Dahn, J. R. Mechanically alloyed Sn-Fe(-C) powders as anode materials for Li-ion batteries. I. The $\text{Sn}_2\text{Fe-C}$ system. *J. Electrochem. Soc.* **146**, 405–413 (1999).
- Idota, Y. et al. Nonaqueous secondary battery. US Patent No. 5,478,671 (1995).
- Sigala, C., Guyomard, D., Piffard, Y. & Tournoux, M. Synthesis and performances of new negative electrode materials for 'Rocking Chair' lithium batteries. *C.R. Acad. Sci. Paris II* **320**, 523–529 (1995).
- Courtney, I. A., McKinnon, W. R. & Dahn, J. R. On the aggregation of tin in SnO composite glasses caused by the reversible reaction with lithium. *J. Electrochem. Soc.* **146**, 59–68 (1999).
- Denis, S., Baudrin, E., Touboul, M. & Tarascon, J.-M. Synthesis and electrochemical properties vs Li of amorphous vanadates of general formula RVO_4 (R = In, Cr, Al, Fe, Y). *J. Electrochem. Soc.* **144**, 4099–4109 (1997).
- Guyomard, D. & Tarascon, J.-M. Rechargeable $\text{Li}_{1+x}\text{Mn}_2\text{O}_4/\text{carbon}$ cells with a new electrolyte composition. Potentiostatic studies and application to practical cells. *J. Electrochem. Soc.* **140**, 3071–3081 (1993).
- Gozdz, A. S., Tarascon, J.-M. & Schmutz, C. N. Rechargeable lithium interaction battery with flexible electrolyte. US Patent No. 5,296,318 (1994).
- Amatucci, G. G., Tarascon, J.-M. & Klein, L. C. CoO_2 , the end member of the Li_xCoO_2 solid solution. *J. Electrochem. Soc.* **143**, 1114–1123 (1996).
- Figlarz, M., Fievet, F. & Lagier, J.-P. Reduction process of metal-based inorganic precursors in liquid polyols to produce monodisperse metal particles. French patent No. 8,221,483 (1985).
- Buffat, P. & Borel, J. P. Size effect on the melting temperature of gold particles. *Phys. Rev. A* **13**, 2287–2292 (1976).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank E. Baudrin, D. Larcher, M. Morcrette, Y. Chabre, G. Amatucci and C. Masquelier for discussions.

Correspondence and requests for materials should be addressed to J.-M.T. (e-mail: Jean-Marie.Tarascon@u-picardie.fr).

Origins of sulphate in Antarctic dry-valley soils as deduced from anomalous ^{17}O compositions

Huiming Bao*, Douglas A. Campbell*, James G. Bockheim† & Mark H. Thiemens*

* Department of Chemistry & Biochemistry, Mail Code 0356, University of California San Diego, 9500 Gilman Drive, La Jolla, California 92093-0356, USA
† Department of Soil Science, University of Wisconsin, 1525 Observatory Drive, Madison, Wisconsin 53706-1299, USA

The dry valleys of Antarctica are some of the oldest terrestrial surfaces on the Earth. Despite much study of soil weathering and development, ecosystem dynamics and the occurrence of life in these extreme environments^{1–3}, the reasons behind the exceptionally high salt content of the dry-valley soils^{4–6} have remained uncertain. In particular, the origins of sulphate are still controversial; proposed sources include wind-blown sea salt^{5,7}, chemical weathering⁸, marine incursion⁹, hydrothermal processes¹⁰ and oxidation of biogenic sulphur in the atmosphere¹. Here we report measurements of $\delta^{18}\text{O}$ and $\delta^{17}\text{O}$ values of sulphates from a range of dry-valley soils. These sulphates all have a large positive anomaly¹¹ of ^{17}O , of up to 3.4‰. This suggests that Antarctic sulphate comes not just from sea salt (which has no anomaly of ^{17}O) but also from the atmospheric oxidation of reduced gaseous sulphur compounds, the only known process that can generate the observed ^{17}O anomaly. This source is more prominent in high inland soils, suggesting that the distributions of sulphate are largely explained by differences in particle size and transport

mode which exist between sea-salt aerosols and aerosols formed from biogenic sulphur emission.

Water-soluble sulphate from soils of four different types of ecologic-climatic regions in Antarctic dry valleys were analysed for $\delta^{18}\text{O}$ and $\delta^{17}\text{O}$ values (see ref. 12 for method). The sulphate $\delta^{18}\text{O}$ and $\delta^{17}\text{O}$ values are among the most negative measured on Earth, ranging from 2.8 to -10.5‰ , and 4.5 to -3.4‰ (all in SMOW), respectively (Table 1). We note the large sulphate ^{17}O anomalies, or mass-independent oxygen isotopic compositions, shown here by the positive $\Delta^{17}\text{O}$ values ($\Delta^{17}\text{O} = \delta^{17}\text{O} - 0.528\delta^{18}\text{O}$), ranging from 0.6 to 3.4‰ . These data also show that the $\Delta^{17}\text{O}$ value increases with increasing depth in the soil profile for nearly all 12 soils analysed (Table 1) and that there exist two groups of data represented by soils from upland valleys and plateau fringe and soils from inland valley floors and sides, with the first showing lower $\delta^{18}\text{O}$ and higher $\Delta^{17}\text{O}$ values than for the second group (Fig. 1).

^{17}O anomalies have only recently been discovered in terrestrial surficial solids^{13,14}. Previously, only meteoritic samples possessed ^{17}O anomalies in the solid regime. Some gas-phase atmospheric molecules are known to possess mass-independent isotopic compositions¹⁵. The extremely positive sulphate $\Delta^{17}\text{O}$ values from Antarctic soils point to an atmospheric origin. A recent study has

shown that the oxidation of reduced sulphur compounds in the atmosphere by O_3 or H_2O_2 is the source of positive $\Delta^{17}\text{O}$ values observed in atmospheric sulphates¹⁶. All other soil sulphate sources, such as seawater or sea-salt sulphate and sulphate from local chemical weathering, possess $\Delta^{17}\text{O} \approx 0$, as shown by direct measurement¹³. In Antarctica, recent anthropogenic SO_2 contamination in the lower troposphere is insignificant¹⁷, overall SO_2 input from volcanoes is minimal, and most reduced sulphur compounds are dominated by biogenic emissions from the ocean, particularly dimethylsulphide (DMS)^{18,19}.

The question arises as to what fraction of the water-soluble sulphate in soil originated from biogenic sulphur emitted from the surrounding oceans. There are two additional sources that could also contribute to sulphates possessing positive $\Delta^{17}\text{O}$: (1) long-range transport of sulphate from continental areas and (2) downward transport of a stratospheric reservoir, which includes a flux of sulphate from the stratosphere through downward intrusion of stratospheric air masses within the Antarctic vortex in winter. These sulphate sources could account for a maximum of 33% in winter and 10% in summer for non-sea-salt (NSS) sulphate in coastal Antarctic regions, a minor component of the total sulphur cycle on an annual basis¹⁹. Unfortunately, no sulphate $\Delta^{17}\text{O}$ data from these

Table 1 Oxygen isotopic compositions for soil sulphates from the dry-valley regions, Antarctica

Sample	Depth (cm)	Na ⁺	Cl ⁻ (mmole l ⁻¹)*	SO ₄ ²⁻	δ ¹⁸ O†	δ ¹⁷ O (‰)	Δ ¹⁷ O	Location
Inland valley floor								
77-1-3	0.5–7	13	4.5	38.0	+1.9	+1.8	+0.83	Wright Valley
77-1-4	7–22	598	437.0	140.0	–0.6	+0.6	+0.95	
84-97-373	0–6	53	35.6	33.0	+0.5	+2.2	+1.98	Wright Valley
84-97-374	6–18	291	151.1	59.8	+1.4	+4.0	+3.24	
83-1-2	0–8	11.5	7.7	22.5	+2.5	+2.0	+0.71	Taylor Valley
83-1-3	8–15	279.7	3.2	293.2	+0.1	+0.6	+0.56	
83-28-122	0–6	32.4	36.3	11.5	–2.6	–0.1	+1.23	Wright Valley
Inland valley side								
79-9-53	0.5–3	163	92.0	71.0	–0.6	+0.9	+1.21	Wright Valley
79-9-54	3–30	608	350.0	95.0	–1.9	+0.9	+1.83	Taylor Valley
83-13-54	0–8	13.6	15.6	14.7	–0.6	+0.5	+0.83	
83-13-55	8–18	95	97.0	16.5	–0.9	+0.4	+0.87	Wright Valley
83-39-159	0–14	17.4	11.1	8.4	+0.1	+1.2	+1.19	
83-41-169	0–14	47.3	32.4	18.8	–1.1	+0.7	+1.29	Wright Valley
83-41-170	14–28	1088.8	747.6	227.3	–2.5	+0.3	+1.64	Wright Valley
83-41-172	70–120	16.2	17.0	17.3	–4.0	–0.5	+1.57	
84-43-172	0–4	144.5	128.7	42.9	+2.0	+2.3	+1.24	Wright Valley
84-43-173	4–16	705.2	748.7	26.3	+1.0	+2.5	+2.00	
84-43-176	63–100	35.5	39.1	5.2	–1.4	+1.5	+2.27	Wright Valley
84-47-192	0–4	21.9	18.2	45.9	–0.6	+0.9	+1.23	
84-47-193	4–20	820.4	947.8	157.6	0.0	+1.9	+1.89	Wright Valley
84-47-195	42–85	30.3	30.6	7.7	+1.7	+2.7	+1.83	
84-49-209	>115	1168	899.8	38.5	+2.8	+4.5	+3.05	Wright Valley
Upland valley								
78-5-34	0.5–5	1340	2.4	710.0	–7.0	–0.9	+2.83	Arena Valley
82-8-25	0–5	74.5	63.7	43.5	–2.1	+0.2	+1.28	Pearse Valley
82-11-37	0–4	4.8	2.3	24.6	–4.5	+0.2	+2.61	Arena Valley
82-11-38	4–25	342	75.9	112.1	–9.8	–2.7	+2.37	
82-11-40	30–100	14.4	2.2	22.6	–6.0	+0.2	+3.38	Arena Valley
82-13-46	0–4	5.6	2.8	35.0	–8.4	–2.4	+1.96	
82-13-47	4–19	163	25.1	164.1	–10.4	–3.4	+2.05	Beacon Valley
82-29-116	4–15	14.8	9.9	40.3	–8.0	–1.9	+2.37	
82-31-123	0–3	8.1	6.6	37.5	–7.6	–1.6	+2.33	Beacon Valley
82-33-129	0–5	17.8	12.4	44.3	–9.5	–2.7	+2.25	Beacon Valley
82-33-131	30–85	12.7	8.4	42.1	–9.1	–2.3	+2.50	
Plateau fringe								
80-18-75	0–6	31	4.2	1.5	–8.8	–2.5	+2.14	South of Mt. Fleming
80-18-76	6–12	345	11.4	34.0	–8.4	–1.8	+2.63	
80-18-78	30–48	30	10.9	3.3	–6.4	0.0	+3.24	

* Ion concentration data are from the Antarctic Soils Laboratory Database, collected by J.G.B. and his students.

† $\delta^{18}\text{O} = [(R_{\text{sample}}/R_{\text{standard}}) - 1] \times 1,000\text{‰}$, where $R = ^{18}\text{O}/^{16}\text{O}$ and the standard is SMOW. The same applies to $\delta^{17}\text{O}$. Mass-dependent fractionation produces a constant relationship of $\delta^{17}\text{O} = 0.528\delta^{18}\text{O}$. A sample with the $\delta^{17}\text{O}/\delta^{18}\text{O}$ pair that deviates from this relationship is said to have a mass-independent isotopic composition.

different Antarctic aerosols are available at this time, which limits a quantitative estimate of the contributions to soil sulphate by differing sources. On the other hand, the Antarctic soil sulphates have nearly the same range of $\Delta^{17}\text{O}$ values as the aerosol sulphates collected in the greater Los Angeles area²⁰. If atmospheric SO_2 oxidation processes in Antarctica are not significantly different from that in the Los Angeles area, a primary source of most soil sulphate in Antarctica (except perhaps in the coastal zone) is sulphate derived from reduced biogenic sulphur, with a source from the surrounding oceans.

The increasing sulphate $\Delta^{17}\text{O}$ value down the soil profile is intriguing. It could be a primary record of changes in sulphate aerosol depositions over time, or simply the result of post-depositional alteration by soil development that has little palaeoclimatic significance. Many of these soils are several million years old²¹, and evidence shows that the sulphur source and budget have changed over time due to changes in climatic conditions. For example, 5 to 20 times higher marine and continental inputs have been observed during glacial times than at present²², and Vostok ice core data show a 20–46% increase in NSS sulphate concentration in the Last Glacial Maximum (after accounting for change in accumulation rate)²³. However, leaching is a continuing process in the whole soil profile, despite the cold and dry conditions, as indicated by the occurrence of salt encrustations in soil profiles^{24,25} and field tests²⁵, as well as the detection of ^{10}Be deep in old soil profiles where the short 1.5-million-year half-life of ^{10}Be renders the original signal undetectable²⁶. A temporal record based upon soil depth has most probably been altered, or even erased. We propose here that the observed increase of $\Delta^{17}\text{O}$ value down a soil profile is the result of preferential leaching of sulphate derived from reduced biogenic sulphur relative to sea-salt sulphate in the soil profile. It is possible that the much smaller size (submicrometres) of NSS sulphate particles will permit preferential dissolution over coarse (micrometres) sea-salt particles. Thus, $\Delta^{17}\text{O}$ -positive sulphate derived from gaseous sulphur oxidation (including stratospheric sources) slowly concentrates at depth. We can also infer that the $\Delta^{17}\text{O}$ value of sulphate derived from biogenic sulphur may be as large as 3.38, equal to the maximum $\Delta^{17}\text{O}$ value detected deep in soil profiles. Similarly, the maximum $\Delta^{17}\text{O}$ value for aerosol sulphate measured in the great Los Angeles area is 3.2 (ref. 20).

An overall smaller sea-salt component in soils of upland valleys compared to valley floors or sides^{1,5,6} is consistent with our $\Delta^{17}\text{O}/\delta^{18}\text{O}$ pattern (Fig. 1). First, the residence time in air for sea salt

(coarse aerosol) is approximately hours to days, significantly shorter than that of the sulphate particles (fine aerosol) derived from reduced biogenic sulphur, which may last a few days to a week²⁷. The sea-salt is deposited adjacent to the coastal area while sulphate derived from biogenic sulphur gets carried further inland and to high altitudes. An enhanced NSS sulphate component in inland high-latitude regions resulted in observed higher $\Delta^{17}\text{O}$ and lower $\delta^{18}\text{O}$ values than in coastal regions, since sea-salt sulphate has a constant $\Delta^{17}\text{O}$ ($\sim 0\text{‰}$) and $\delta^{18}\text{O}$ ($\sim 9.4\text{‰}$).

Second, the correlation between $\Delta^{17}\text{O}/\delta^{18}\text{O}$ and climatic region is also related to previous observations that the origin of precipitation largely determines the ion content in soils of different regions: drifting snow from the central plateau for the upland and inland regions, and local moisture with a high content of sea salt from the Ross Sea for the coastal regions^{1,5,6}. Overall the amount of biogenic sulphur in Antarctica may vary with climate (including season) and the extent of sea ice²⁸.

The Antarctic dry-valley regions are the oldest terrestrial surfaces²¹. It is interesting to compare the results from this region with an equally old (since the Miocene period) surface, the central Namib desert, where the massive gypcrete deposits also have positive sulphate ^{17}O anomalies, but of much smaller magnitude, with most $\Delta^{17}\text{O}$ at 0.3–0.4 (ref. 13). Several factors may contribute to the lower $\Delta^{17}\text{O}$ values of the Namib: for example, low elevation and proximity to the ocean, resulting in a high flux of sea salt onto the land; higher temperature facilitating weathering and microbial activities, and perhaps more sulphate input from parental rocks. It is also important to note that the dry-valley soil data are a composite record over the past several millions of years. The overall input of volcanic sulphur gases, which can also contribute $\Delta^{17}\text{O}$ -positive sulphate to the dry valleys upon oxidation in the atmosphere, although small, is not known. Future measurements of multiple oxygen isotopes in soil nitrate (believed to be largely derived from the stratosphere²⁹) as well as of multiple sulphur isotopes in sulphate, will provide further constraints on the origin, transport, mixing, and post-depositional alterations of the salt deposits in Antarctic dry valleys. The $\delta^{34}\text{S}$ of sulphate alone is limited in distinguishing between biogenic and seawater sulphate owing to their similar range, with the former about 13–22‰ (CDT) and the latter about 21‰. Simultaneous measurement of sulphate $\delta^{18}\text{O}$ and $\Delta^{17}\text{O}$ could better address the origin of other types of sulphate in Antarctica, for example, the thick-bedded sulphate salt deposits found in coastal regions¹. In addition, mass-independent oxygen and sulphur isotopic compositions have been observed in martian meteoritic secondary minerals³⁰. The large positive $\Delta^{17}\text{O}$ values from Antarctic soil sulphates added another conspicuous similarity between martian surface environments and the Antarctic dry valleys. It also suggests a possibility that anomalous isotopic signatures detected on meteorites could come from terrestrial contamination.

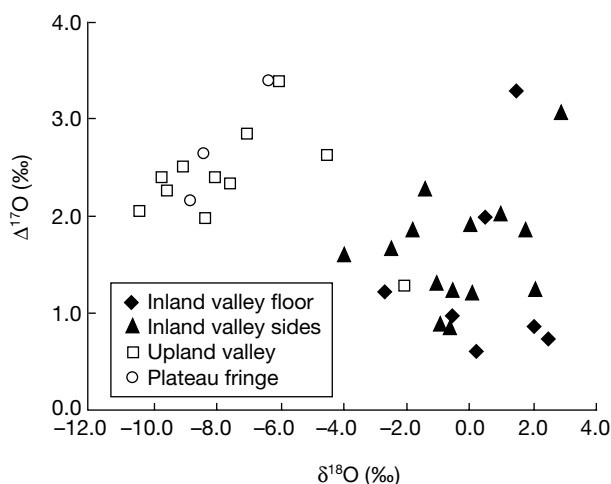


Figure 1 Sulphate $\delta^{18}\text{O}$ and $\Delta^{17}\text{O}$ values for soils from different climatic regions in Antarctic dry valleys. The error bar for $\delta^{18}\text{O}$ value can be as large as ± 0.8 (not shown); it is, however, only ± 0.05 for the $\Delta^{17}\text{O}$ value, owing to the covariations of $\delta^{18}\text{O}$ and $\delta^{17}\text{O}$ during measurement¹².

Received 2 May; accepted 14 August 2000.

- Campbell, I. B. & Claridge, G. G. C. *Antarctica: Soils, Weathering Processes And Environment* (Elsevier, Amsterdam, 1987).
- Priscu, J. C. *Ecosystem Dynamics In A Polar Desert; The McMurdo Dry Valleys, Antarctica* (American Geophysical Union, Washington DC, 1998).
- Gibson, E. K., Wentworth, S. J. & McKay, D. S. Chemical weathering and diagenesis of a cold desert soil from Wright Valley, Antarctica: an analog of Martian weathering processes. *J. Geophys. Res.* **A88**, 912–928 (1983).
- Claridge, G. G. C. & Campbell, I. B. The salts in Antarctic soils, their distribution and relationship to soil processes. *Soil Sci.* **123**, 377–384 (1977).
- Key, J. R. & Williams, K. Origin of crystalline, cold desert salts in the McMurdo region, Antarctica. *Geochim. Cosmochim. Acta* **45**, 2299–2309 (1981).
- Bockheim, J. G. Properties and classification of cold desert soils from Antarctica. *Soil Sci. Soc. Am. J.* **61**, 224–231 (1997).
- Torii, T., Yamagata, N., Osaka, J. & Murata, S. A view on the formation of saline waters in the Dry Valleys. *Mem. Natl Inst. Polar Res., Spec. Issue (Jpn)* **13**, 22–33 (1979).
- Linkletter, G. O. Weathering and soil formation in the dry valleys of southern Victoria Land. Possible origin for the salts in the soils. *Antarct. Geol. Geophys., Symp. Antarct. Geol. Solid Earth Geophys. B* **1**, 441–446 (1972).
- Nakai, N., Kiyosu, Y., Wada, H. & Takimoto, M. Stable isotope studies of salts and water from Dry

- Valleys, Antarctica. I. Origin of salts and water, and the geologic history of Lake Vanda. *Mem. Natl. Inst. Polar Res., Spec. Issue (Jpn)* **4**, 30–44 (1975).
10. Takamatsu, N., Kato, N., Matsumoto, G. I. & Torii, T. Salt origin viewed from lithium distributions in lake and pond waters in the McMurdo Dry Valleys, Antarctica. *Verh. Int. Verein. Theor. Angew. Limnol.* **25**, 954–956 (1993).
 11. Clayton, R. N., Grossman, L. & Mayeda, T. K. A component of primitive nuclear composition in carbonaceous chondrites. *Science* **182**, 485–488 (1973).
 12. Bao, H. & Thiemens, M. H. Generation of O_2 from $BaSO_4$ using a CO_2 -laser fluorination system for simultaneous analysis of $\delta^{18}O$ and $\delta^{17}O$. *Anal. Chem.* **72**, 4029–4032 (2000).
 13. Bao, H. *et al.* Anomalous ^{17}O compositions in massive sulphate deposits on the Earth. *Nature* **406**, 176–178 (2000).
 14. Lee, C. C. -W. & Thiemens, M. H. $\delta^{17}O$ and $\delta^{18}O$ measurements of atmospheric sulfate from a coastal and high alpine regions: A mass independent isotopic anomaly. *J. Geophys. Res.* (submitted).
 15. Thiemens, M. H. Atmosphere science- Mass-independent isotope effects in planetary atmospheres and the early solar system. *Science* **283**, 341–345 (1999).
 16. Savarino, J., Lee, C. -W. C. & Thiemens, M. H. Laboratory oxygen isotopic study of atmospheric $S(IV)$ oxidation: origin of the mass independent oxygen isotopic anomaly in atmospheric sulfates and other sulfate deposits. *J. Geophys. Res.* (in the press).
 17. Legrand, M. Sulphur-derived species in polar ice; a review. *Ice Core Studies Glob. Biogeochem. Cycles* **30**, 91–119 (1995).
 18. Prospero, J. M., Savoie, D. L., Saltzman, E. S. & Larsen, R. Impact of oceanic sources of biogenic sulphur on sulphate aerosol concentrations at Mawson, Antarctica. *Nature* **350**, 221–223 (1991).
 19. Minikin, A. *et al.* Sulfur-containing species (sulfate and methanesulfonate) in coastal Antarctic aerosol and precipitation. *J. Geophys. Res.* **103**, 10975–10990 (1998).
 20. Bao, H., Michalski, G. & Thiemens, M. H. Sulfate oxygen-17 anomalies in desert varnishes. *Geochim. Cosmochim. Acta* (in the press).
 21. Schafer, J. *et al.* Cosmogenic noble gas studies in the oldest landscape on earth: surface exposure ages of the Dry Valleys, Antarctica. *Earth Planet. Sci. Lett.* **167**, 215–226 (1999).
 22. Petit, J. R., Briat, M. & Royer, A. Ice age aerosol content from East Antarctic ice core samples and past wind strength. *Nature* **293**, 391–394 (1981).
 23. Cunningham, J. & Waddington, E. D. Air flow and dry deposition of non-sea salt sulfate in Polar firm: paleoclimatic implications. *Atmos. Environ.* **A 27**, 2943–2956 (1993).
 24. Bockheim, J. G. Relative age and origin of soils in eastern Wright Valley, Antarctica. *Soil Sci.* **128**, 142–153 (1979).
 25. Claridge, G. G., Campbell, I. B. & Balks, M. R. in *Ecosystem Processes In Antarctic Ice-Free Landscapes* (eds Howard-Williams, C. & Hawes, I.) 137–143 (1997).
 26. Graham, I. J., Ditchburn, R. G., Sparks, R. J. & Whitehead, N. E. ^{10}Be investigations of sediments, soils and loess at GNS. *Nucl. Instrum. Methods. Phys. Res. B* **123**, 307–318 (1997).
 27. Seinfeld, J. H. & Pandis, S. N. *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change* (Wiley, New York, 1998).
 28. Welch, K. A., Mayewski, P. A., Whitlow, S. I. Methanesulfonic acid in coastal Antarctic snow related to sea-ice extent. *Geophys. Res. Lett.* **20**, 443–446 (1993).
 29. Wagenbach, D., Legrand, M., Fischer, H., Pichlmayer, F. & Wolff, E. W. Atmospheric near-surface nitrate at coastal Antarctic sites. *J. Geophys. Res.* **A 103**, 11007–11020 (1998).
 30. Farquhar, J., Savarino, J., Jackson, T. L. & Thiemens, M. H. Evidence of atmospheric sulphur in the Martian regolith from sulphur isotopes in meteorites. *Nature* **404**, 50–52 (2000).

Acknowledgements

We thank D. S. Sheppard for discussion on his recent isotopic data for sulphate and nitrate in Antarctic soils, and J. Savarino for comments. We thank NASA and the NSF for support.

Correspondence and requests for materials should be addressed to H.B. (e-mail: hbao@chem.ucsd.edu).

Transient high temperatures in mantle plume heads inferred from magnesian olivines in Phanerozoic picrites

R. N. Thompson* & S. A. Gibson†

* Department of Geological Sciences, University of Durham, South Road, Durham DH1 3LE, UK

† Department of Earth Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EQ, UK

Both scaled laboratory experiments and numerical models of terrestrial mantle plumes produce ‘balloon-on-a-string’ structures, with a bulbous head followed by a stem-like tail. Discussions have focused on whether their initial upwelling heads are hotter than the tails or cooler, as a result of entrainment of

ambient mantle during ascent^{1–3}, and also on whether initial plume upwelling is a newtonian or non-newtonian process^{4,5}. The temperature of the mantle delivered to the base of the lithosphere is a critical parameter in such debates. Dry continental magmas can normally contribute little to this topic because their hottest (ultramafic) examples can be expected to be trapped, owing to their density, beneath the Moho. Here we report a rare case in which olivine (with 93.3% forsterite; Mg_2SiO_4) phenocrysts, precipitated from an unerupted komatiitic melt (~24% MgO) of the Tristan mantle plume head 132 Myr ago, were carried to upper-crust levels in northwest Namibia by less Mg-rich (9.6–18.5% MgO) magmas. We infer that the hidden melt, generated when the plume impinged on the base of the lithosphere, originated in the mantle with a potential temperature of ~1,700 °C. This is ~400 °C above ambient and much hotter than the temperatures previously calculated for steady-state Phanerozoic mantle plumes^{3,6–8}. Published data show that the same conclusion can be reached for the initial Iceland and Galapagos plumes.

All currently active mantle plumes (for example, Hawaii, Iceland, Galapagos, Tristan, Réunion) have been present, judging from their associated magmatism, for several tens of millions of years. The temperatures within these plumes may be estimated independently from geophysical observations and the geochemistry of their melts (for example, ref. 3). For Hawaii and Iceland, diverse approaches have produced concordant estimates^{3,6–8} that the potential temperature (T_p ; mantle temperature projected along the solid-state adiabat to one atmosphere) in these steady-state plumes is ~150–300 °C above that of the global asthenospheric MORB-source upper mantle (that is, $T_p = 1,280$ – $1,350$ °C for MORB—mid-ocean-ridge basalt—and $1,450$ – $1,600$ °C for Hawaii and Iceland). Because no new plumes are at present impinging on the Earth’s lithosphere, the temperatures of starting-plume heads can only be investigated by models (experimental or numerical) or via the volume, eruption rate and geochemistry of melts from former examples. White and McKenzie³ found no evidence from rare-earth element (REE) inversion studies that starting-plume magmas had hotter mantle

Table 1 Chemical analyses of Horingbaai dykes, their olivine and calculated parental magma

	1	2	3†	4	5
SiO ₂	47.27	47.64	46.4	40.24	40.93
TiO ₂	1.02	1.13	0.7		
Al ₂ O ₃	13.28	14.01	9.0		
Cr ₂ O ₃				0.12	0.13
Fe total*	11.86	12.62	12.0	12.02	6.65
NiO				0.25	0.34
MnO	0.17	0.18	0.2	0.19	0.05
MgO	13.56	10.37	24.0	46.81	51.73
CaO	10.90	11.86	6.5	0.37	0.27
Na ₂ O	1.55	1.94	1.1		
K ₂ O	0.30	0.21	0.1		
P ₂ O ₅	0.09	0.1	0.1		
Total	100.00	100.06	100.0	100.08	100.10
LOI	0.76	0.42			
Mg#	76.9	65.4	81.3	87.4	93.3

Analyses of rocks on ignited powders. LOI, loss on ignition. See ref. 30 for methods. Olivine analyses were by electron microprobe, using both wavelength- (WDS) and energy-dispersive (EDS) methods. EDS operating conditions: 30 nA beam current; 180 s count time; average of 3 spot analyses. EDS reproducibility (2σ), based on repeated analyses of an olivine standard: SiO₂ = 0.3; MgO = 0.35; FeO = 0.2; other oxides = 0.1 wt%. WDS operating conditions (for Ca, Cr, Ni): 30 nA beam current; 100 s count time; average of 3 spot analyses. WDS reproducibility (2σ): Ca 0.008; Cr 0.016; Ni 0.02. Here we use the common practice (for example, refs 11, 22) of referring in the text to olivine Mg# as per cent forsterite (Fo). All data used in construction of Fig. 1b have been recalculated in the same way. Columns have meanings as follows. 1. Dyke 97SB33 (21° 18′ 35″ S, 14° 42′ 43″ E). This dyke contains the largest proportion of Fo₉₃ olivines, about half of its macrocryst content. 2. Dyke 97SB29 (21° 25′ 33″ S, 13° 53′ 49″ E). This dyke is approximately the composition in equilibrium with the Horingbaai dyke euhedral olivine macrocrysts. 3. Estimated composition of the magma in equilibrium with the Fo₉₃ macrocrysts. This is strictly a komatiite¹⁴. 4. Euhedral olivine macrocryst in 97SB33. 5. Subhedral olivine macrocryst in 97SB33.

* All Fe as Fe₂O₃ in rocks and FeO in olivines. Mg# in rocks and parental magma calculated with 9.5% of total Fe as Fe₂O₃ (see Table 2, note ‡).

† Fe₂O₃ calculated from MgO, to maintain Mg#. Other oxides estimated from oxide-MgO plots of the Horingbaai suite.

sources than those of steady-state plumes. Nevertheless, they suggested that this might be so, in order to explain the fall in melt production rate from plumes with time after they reached the lithosphere. Griffith and Campbell¹ argued, from laboratory convection experiments, that large-scale thermal entrainment and mixing with ambient mantle during the initial ascent of a plume head would cool it, relative to the associated tail. Farnetani and Richards² suggested, from numerical modelling using newtonian rheology, that such thermal entrainment is a less important process, and that the excess initial plume-head temperature is 300–350 °C (that is, potential temperature $T_p \approx 1,600$ – $1,650$ °C). Here we present petrological evidence pointing to high plume-head temperatures.

Studies using the melt products of experimental phase equilibria to calculate melt compositions produced during mantle decom-

pression (for example, ref. 9) have led to attempts to calculate plume-head T_p from the compositions of Mg-rich magmas. Thus, Gill *et al.*¹⁰ calculated $T_p = 1,550$ – $1,600$ °C (or possibly $1,650$ °C, depending on the value of entropy of melting used) for the head of the Iceland plume when it initially impinged on the lithosphere beneath West Greenland/Baffin Bay about 61 Myr ago. They equated the most forsteritic olivine phenocrysts in West Greenland picrites (with 92.3% forsterite; $Fo_{92.3}$) to a liquid with 19% MgO. Nisbet *et al.*¹¹ used the same approach with komatiites that have preserved fresh olivine: Alexo, $Fo_{94.1}$, MgO = 28.1%; Belingwe, $Fo_{93.6}$, MgO = 25.7%; Gorgona, $Fo_{91.0}$, MgO = 18.0. Nisbet *et al.* calculated $T_p = 1,800$ °C for the 2.7-Gyr-old Belingwe and Alexo komatiite sources (deduced by them to be plumes), and $1,600$ °C for Gorgona at 87 Myr (generally considered to be a product of the initial Galapagos plume head).

Two Phanerozoic examples have been reported of Mg-rich magmas containing phenocrysts/xenocrysts (magmatic, not mantle) of olivine $\sim Fo_{93}$ (Fig. 1). In Gorgona “picrite” dykes (strictly komatiites; see IUGS definitions below) the olivines are euhedral phenocrysts¹². In Padloping island lavas, Baffin Bay, they are xenocrysts, but Francis¹³ argued that these had been phenocrysts in magmas that had failed to reach the surface. L. M. Larsen (personal communication) has recently found euhedral-subhedral Fo_{93} macrocrysts in other West Greenland picrites. Here we report Fo_{93} olivines in basalt-picrite dykes (Table 1) of the Horingbaai suite, northwest Namibia, in the southeast part of the Paraná-Etendeka large igneous province. We use the composition boundaries for melts of 12% MgO between basalt and picrite and 18% between picrite and komatiite, as recently agreed by the IUGS Subcommittee on the Systematics of Igneous Rocks¹⁴. Published accounts of the type-locality coastal Horingbaai dolerites have emphasised their MORB-like geochemistry: 7–9.5% MgO; flat

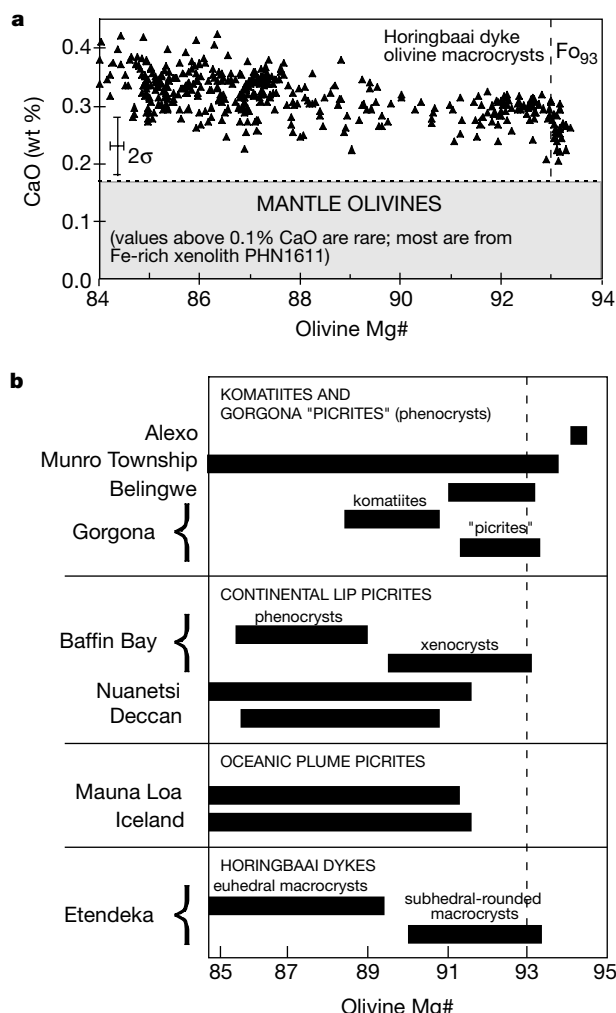


Figure 1 Affinities of the Horingbaai dyke olivines. The plot of olivine CaO as a function of Mg# (% forsterite) (a) shows that the macrocrysts crystallized from a related suite of magmas and are not xenocrysts from mantle or other sources (see text). Note the relative lack of analyses at $\sim Fo_{90}$; a reflection of the bimodality of olivine compositions in some of the dykes. Panel (b) compares the Mg# (100Mg/(Mg + Fe)) ranges of the Horingbaai phenocrysts and xenocrysts with those of olivines in continental and oceanic picrites associated with mantle plumes, and with Archaean komatiites. LIP, large igneous province. The Baffin Bay suite contains more Mg-rich olivines than other picrites of the 63–60 Myr North Atlantic LIP (for example, East Greenland, Skye). Other continental LIP picrite suites (for example, Keweenaw) have either olivine less Mg-rich than Fo_{90} or entirely altered olivine. See text for discussion of the terms picrite and komatiite, and a note on the Siberian meymechites. For a list of the data sources for this figure, see Supplementary Information.

Table 2 Calculations of Mg# and MgO content of the magma that precipitated $Fo_{93.3}$ phenocrysts

Assumed depth of olivine precipitation*	“Near-surface”	Base of thinned crust	Base of thick Damara belt crust
$ K_d^{Fe^{2+}/Fe^{3+}}$ (ref. 19)†	~ 1 atm	~ 20 km; 0.5 GPa	~ 50 km; 1.2 GPa
Mg# of magma	0.30	0.31	0.33
MgO of magma‡	80.8	81.3	82.2
Calculated plume T_p	23.3	24.0	25.5
		$\sim 1,680$ °C	

* See text. Our preferred model is olivine precipitation at the base of thinned crust offshore, followed by lateral dyke injection. The other models are limiting cases. Thus, the “near-surface” model treats the dykes as lavas and the “thick-crust” scenario makes the implausible assumption that the MORB-like Horingbaai magmas were all generated beneath thick Damara belt lithosphere and then migrated laterally towards the adjacent zone of pre-Atlantic rifting, rather than vice versa.

† Ulmer¹⁹ quotes 2σ errors of ± 0.015 on these K_d values.

‡ Calculated from a plot of MgO versus Mg# for our own Horingbaai dyke analyses, with Fe_2O_3 set between 0.07 and 0.12 of total Fe (preferred 0.095). These values are the full range of direct Fe^{2+}/Fe^{3+} measurements for unoxidized Hawaiian tholeiitic-transitional basalt glasses³¹. We also attempted to determine more directly the oxidation state of the magmas that precipitated the Horingbaai olivine macrocrysts, by microprobe analyses of Cr spinels and the olivines enclosing them, and applying the geothermometer of Sack and Ghiorso³². This thermodynamic model was preferred because it is the most successful of several published since 1990 in reproducing the f_{O_2} range for Hawaiian magmas that is implied by the Fe^{3+}/Fe^{2+} of their glasses. Horingbaai macrocryst olivine-spinel equilibration temperatures are all $< 1,020$ °C, with corresponding f_{O_2} between 4 and 8 log units below the fayalite = magnetite + quartz (FMQ) buffer; clearly products of local sub-solidus re-equilibration.

§ Our estimated MgO content for the magma that precipitated the $Fo_{93.3}$ phenocrysts. Errors were calculated as follows: the 2σ ranges for olivine microprobe Mg and Fe analysis (Table 1) and for K_d (†) were propagated to give minimum and maximum values of 80.0 and 82.6 for magma Mg#. Combining these with the Fe_2O_3 range gave three graphs to determine MgO, each fitted by a polynomial curve with $R^2 > 0.99$. The resulting magma MgO values are: 24.0% preferred; 21.5% minimum; 26.6% maximum.

|| Method of ref. 9; see text. The equation of Renner (in ref. 11) was used to calculate preferred (1,477 °C), minimum (1,432 °C) and maximum (1,523 °C) 1-atm liquid from the MgO range, using the 1-atm model above. After pressure correction to 50 km depth (see text), these were plotted on a T - X section at 1.5 GPa for KLB-1. The section incorporated data from both Takahashi *et al.*²⁹ and Falloon *et al.*³³, who studied the closely similar composition MM3. The polynomial curve fitting both data sets ($R^2 > 0.99$) was used to calculate the following values for melt fraction (X): preferred 0.44; minimum 0.37; maximum 0.49. The preferred value differs from the median because of the sigmoidal shape of the T - X curve (Fig. 2). Published detailed studies of KLB-1 and MM3 phase equilibria since 1993^{25,26,33} locate the solidus to within ± 25 °C at pressures below 7 GPa. Combined with the ranges of T and X at 1.5 GPa, given above, the calculated values of T_p are: preferred 1,680 °C (6.0 GPa); minimum 1,630 °C (5.0 GPa); maximum 1,770 °C (7.5 GPa); corresponding intersection pressures with the KLB-1 solidus in brackets. The range of T_p can be increased by taking more pessimistic estimates of the error associated with the KLB-1 solidus but phase-equilibria studies since 1993^{25,26,33} do not support this.

Olivine macrocrysts in the dykes range up to 6 mm in size

Here we focus on the more Mg-rich olivines and their impli-

Although variable, Cr_2O_3 tends to be higher in the Horingbaai

The Mg# of the olivine forsteritic macrocrysts can be used to

We therefore deduce that the most Mg-rich olivine in the

have formed in melts about 100 °C hotter than those that became the spinifex-textured komatiites; likewise the Baffin Bay magmas that precipitated olivine with $Fo > 93$. The only way to disturb these thermal relationships is to postulate that some of the magmas were hydrous and others were not, but we have found no traces of hydrous minerals in the interstices of coarsely holocrystalline Horingbaai dykes. In this respect they contrast with the meymechites of the Siberian large igneous province, which also have ~25% MgO and Fo_{93} olivine, but with common interstitial hydroxy-biotite²⁴.

Assuming homogeneous stretching, the thinned crust at ~132 Myr above our preferred site of Horingbaai komatiite separation indicates a lithospheric thickness (to the base of the mechanical boundary layer⁹ above convecting mantle) of ~50 km, as White and McKenzie³ also found by REE inversion modelling of the genesis of the Horingbaai suite. The pressure-corrected komatiite liquidus is estimated to be 1,527 °C, using an adiabatic gradient for the melt of 1 °C km⁻¹ (ref. 9). Plotting this point (filled square) on a pressure–temperature (P – T) diagram for peridotite KLB-1 mantle^{25,26} gives an estimated melt fraction for the 24.0% MgO komatiite of melt fraction $X = 0.44$ (Fig. 2). Finally, the value of T_p of KLB-1 mantle required to generate this melt fraction at 50 km depth by anhydrous decompression was calculated by the method of ref. 9, taking 350 J kg⁻¹ K⁻¹ as the entropy of fusion (ref. 11; M. J. Bickle, personal communication). The result is $T_p = 1,680$ °C (Fig. 2), with dry melting beginning at 6 GPa (~180 km depth).

The substantial amount of total accumulated errors in this calculation (Table 2) means that the absolute value of T_p is less secure than its difference from T_p values in other magmatic suites, calculated the same way. We show this by using the KLB-1 P – T diagram approach for the 17.5% MgO Mauna Loa picrite^{22,23} which gives $X = 0.18$ at 1,487 °C at 72 km depth (base of the sub-Hawaiian mechanical boundary layer⁶), between Watson and McKenzie's mean and maximum values of $X = 0.07$ and 0.22, respectively. The calculated Hawaiian plume T_p from this input is 1,525 °C (Fig. 2), compared with published values^{6,7} of 1,500–1,600 °C.

The tectonomagmatic feature that the 24–25% MgO komatiites of Horingbaai, Gorgona and Baffin Bay have in common is that they were all generated during the initial impinging of Phanerozoic mantle plumes: Tristan, Galapagos and Iceland. The mantle potential temperature that these Mg-rich melts imply (~1,700 °C) is 100–250 °C more than the present-day values for the Hawaii and Iceland plumes^{3,6–8}, and ~400 °C above ambient upper mantle. It is also ~100 °C higher than published T_p estimates for less Mg-rich West Greenland picrites¹⁰ and the Gorgona spinifex-textured komatiites¹¹. This is evidence that the core of a rising mantle plume head contains a fraction of peridotite that is considerably hotter than any that upwells later, as White and McKenzie³ suspected but could not prove directly. Only a small proportion of a plume head can be extremely hot because otherwise implausible volumes of decompression melt will be generated^{13,6,7,9} and these are not seen in the seismic images of the products of Paraná-Etendeka magmatism²⁰. Very hot mantle within the core of a plume head will cool conductively during initial upwelling, and its preservation at the base of the lithosphere is consistent with fluid dynamical models that predict extremely fast non-newtonian ascent^{4,5}. Clearly, despite the “MORB-like” geochemical nature of the Horingbaai magmas, their mantle source was so hot that it must have been within the axial core of a plume impinging on the lithosphere, although non-newtonian plume heads do not have simple cores and margins^{4,5}. Gill *et al.*¹⁰ and Arndt *et al.*²⁷ have discussed, for West Greenland/Baffin Bay and Gorgona respectively, how such “MORB-like” magmas could be generated by decompression melting that starts very deep in a mantle plume beneath a relatively thin lithospheric lid (compare Fig. 2).

Our results assist the debate as to whether the “MORB-source-like” geochemical component of mantle plumes originates within

the D layer or is entrained during plume ascent^{28,29}. If the source mantle of the “MORB-like” melts in the Tristan, Galapagos and Iceland starting plumes reached $T_p \approx 1,700$ °C, such a temperature is too high to be consistent with entrained material—pointing instead to a D" origin. □

Received 4 October 1999; accepted 12 July 2000.

- Griffiths, R. W. & Campbell, I. H. Stirring and structure in mantle starting plumes. *Earth Planet. Sci. Lett.* **99**, 66–78 (1990).
- Farnetani, C. G. & Richards, M. A. Numerical investigation of the mantle plume initiation model for flood basalt events. *J. Geophys. Res.* **99**, 13813–13833 (1994); Thermal entrainment and melting in mantle plumes. *Earth Planet. Sci. Lett.* **136**, 251–267 (1995).
- White, R. S. & McKenzie, D. Mantle plumes and flood basalts. *J. Geophys. Res.* **100**, 17543–17585 (1995).
- van Keken, P. Evolution of starting mantle plumes: a comparison between numerical and laboratory models. *Earth Planet. Sci. Lett.* **148**, 1–11 (1997).
- Larsen, T. B., Yuen, D. A. & Storey, M. Ultrafast mantle plumes and implications for flood basalt volcanism in the North Atlantic Region. *Tectonophysics* **311**, 31–43 (1999).
- Watson, S. & McKenzie, D. Melt generation by plumes: a study of Hawaiian volcanism. *J. Petrol.* **32**, 501–537 (1991).
- Ribe, N. M. & Christensen, U. R. The dynamical origin of Hawaiian volcanism. *Earth Planet. Sci. Lett.* **171**, 517–531 (1999).
- Shen, Y., Solomon, S. C., Bjarnason, I. Th. & Wolfe, C. J. Seismic evidence for a lower-mantle origin of the Icelandic plume. *Nature* **395**, 62–65 (1998).
- McKenzie, D. & Bickle, M. J. The volume and composition of melt generated by extension of the lithosphere. *J. Petrol.* **29**, 625–679 (1988).
- Gill, R. C. O., Pedersen, A. K. & Larsen, J. G. in *Magmatism and the Causes of Continental Break-up* (eds Storey, B. C., Alabaster, T. & Pankhurst, R. J.) 335–348 (Spec. Publ. 68, Geological Society, London, 1992).
- Nisbet, E. G., Cheadle, M. J., Arndt, N. T. & Bickle, M. J. Constraining the potential temperature of the Archaean mantle: a review of the evidence from komatiites. *Lithos* **30**, 291–307 (1993).
- Echeverria, L. M. & Aitken, B. G. Pyroclastic rocks: another manifestation of ultramafic volcanism on Gorgona island, Colombia. *Contrib. Mineral. Petrol.* **92**, 428–436 (1986).
- Francis, D. The Baffin Bay lavas and the value of picrites as analogues of primary magmas. *Contrib. Mineral. Petrol.* **89**, 144–154 (1985).
- Le Bas, M. J. IUGS reclassification of the high-Mg and picritic volcanic rocks. *J. Petrol.* (in the press).
- Duncan, A. R., Armstrong, R. A., Erlank, A. J., Marsh, J. S. & Watkins, R. T. in *Mafic Dykes and Emplacement Mechanisms* (eds Parker, A. J., Rickwood, P. C. & Tucker, D. H.) 119–129 (Balkema, Rotterdam, 1990).
- Schmitt, A. K., Emmermann, R., Trumbull, R. B., Böhn, B. & Henjes-Kunst, F. Petrogenesis and ⁴⁰Ar/³⁹Ar geochronology of the Brandberg complex, Namibia: evidence for a major mantle contribution in metaluminous and peralkaline granites. *J. Petrol.* **41**, 1207–1239 (2000).
- Libourel, G. Systematics of calcium partitioning between olivine and silicate melt: implications for melt structure and calcium content of magmatic olivines. *Contrib. Mineral. Petrol.* **136**, 63–80 (1999).
- Li, J.-P., O'Neill, H. St. C. & Seifert, F. Subsolidus phase relations in the system MgO–SiO₂–Cr–O in equilibrium with metallic Cr, and their significance for the petrochemistry of chromium. *J. Petrol.* **36**, 107–132 (1995).
- Ulmer, P. The dependence of the Fe²⁺–Mg cation-partitioning between olivine and basaltic liquid on pressure, temperature and composition. *Contrib. Mineral. Petrol.* **101**, 261–273 (1989).
- Gladczenko, T. P. *et al.* South Atlantic volcanic margins. *J. Geol. Soc. Lond.* **154**, 465–470 (1997).
- Miller, R. McG. in *Evolution of the Damara Orogen of South West Africa/Namibia* (ed. Miller, R. McG.) 431–515 (Spec. Publ. 11, Geological Society of South Africa, 1983).
- Garcia, M. O., Hulsebosch, T. P. & Rhodes, J. M. in *Mauna Loa Revealed: Structure, Composition, History and Hazards* (eds Rhodes, J. M. & Lockwood, J. P.) 219–239 (Geophys. Monogr. 92, American Geophysical Union, Washington DC, 1995).
- Monierth, C., Johnston, A. D. & Cashman, K. V. in *Mauna Loa Revealed: Structure, Composition, History and Hazards* (eds Rhodes, J. M. & Lockwood, J. P.) 207–217 (Geophys. Monogr. 92, American Geophysical Union, Washington DC, 1995).
- Arndt, N., Chauvel, C., Czamanske, G. & Fedorenko, V. Two mantle sources, two plumbing systems; tholeiitic and alkaline magmatism of the Maymecha River basin, Siberian flood volcanic province. *Contrib. Mineral. Petrol.* **133**, 297–313 (1998).
- Takahashi, E., Shimazaki, T., Tsuzaki, Y. & Yoshida, H. Melting study of a peridotite KLB-1 to 6.5 GPa and the origin of basaltic magmas. *Philos. Trans. R. Soc. Lond. A* **342**, 105–120 (1993).
- Herzberg, C. & Zhang, J. Melting experiments on anhydrous peridotite KLB-1: compositions of magmas in the upper mantle and transition zone. *J. Geophys. Res.* **101**, 8271–8295 (1996).
- Arndt, N. T., Kerr, A. C. & Tarney, J. Dynamic melting in plume heads: the formation of Gorgona komatiites and basalts. *Earth Planet. Sci. Lett.* **146**, 289–301 (1997).
- Hards, V. L., Kempton, P. D. & Thompson, R. N. The heterogeneous Iceland plume: new insights from the alkaline basalts of Snaefell volcanic centre. *J. Geol. Soc. Lond.* **152**, 1003–1009 (1995).
- Hauri, E. H. Major-element variability in the Hawaiian mantle plume. *Nature* **382**, 415–419 (1996).
- Gibson, S. A., Thompson, R. N. & Dickinson, A. P. Ferropicrites: geochemical evidence for Fe-rich streaks in upwelling mantle plumes. *Earth Planet. Sci. Lett.* **174**, 355–374 (2000).
- Byers, C., Garcia, M. & Muenow, D. Volatiles in pillow rim glasses from Loihi and Kilauea volcanoes, Hawaii. *Geochim. Cosmochim. Acta* **49**, 1887–1896 (1985).
- Sack, R. O. & Ghiorso, M. S. Chromian spinels as petrogenetic indicators: thermodynamic and petrological applications. *Am. Mineral.* **76**, 827–847 (1991).
- Falloon, T. J., Green, D. H., Danyushevsky, L. V. & Faul, U. H. Peridotite melting at 1.0 and 1.5 GPa: an experimental evaluation of techniques using diamond aggregates and mineral mixes for determination of near-solidus melts. *J. Petrol.* **40**, 1343–1375 (1999).

Supplementary Information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank N. Arndt, M. Bickle, R. Hardy, D. Jerram, G. Milne, S. Milner, A.-K. Ngono, S. Reed, P. M. Smith and M. Tucker for assistance and discussions, and M. Garcia, R. Gill, G. Pearson and P. M. Smith for comments that substantially improved the manuscript. Durham and Cambridge universities funded the research in part.

Correspondence and requests for materials should be addressed to R.N.T. (e-mail: r.n.thompson@durham.ac.uk).

Evolutionary instability of ectomycorrhizal symbioses in basidiomycetes

David S. Hibbett*†, Luz-Beatriz Gilbert*† & Michael J. Donoghue*†

*Harvard University Herbaria, 22 Divinity Avenue, Cambridge, Massachusetts 02138, USA

Mycorrhizae, the symbiotic associations of plant roots and fungal hyphae, are classic examples of mutualisms. In these ecologically important associations, the fungi derive photosynthetic sugars from their plant hosts, which in turn benefit from fungus-mediated uptake of mineral nutrients. Early views on the evolution of symbioses suggested that all long-term, intimate associations tend to evolve toward mutualism. Following this principle, it has been suggested that mycorrhizal symbioses are the stable derivatives of ancestral antagonistic interactions involving plant parasitic fungi¹. Alternatively, mutualisms have been interpreted as inherently unstable reciprocal parasitisms, which can be disrupted by conflicts of interest among the partners^{2–5}. To determine the number of origins of mycorrhizae, and to assess their evolutionary stability, it is necessary to understand the phylogenetic relationships of the taxa involved. Here we present a broad phylogenetic analysis of mycorrhizal and free-living homobasidiomycetes (mushroom-forming fungi). Our results indicate that mycorrhizal symbionts with diverse plant hosts have evolved repeatedly from saprotrophic precursors, but also that there have been multiple reversals to a free-living condition. These findings suggest that mycorrhizae are unstable, evolutionarily dynamic associations.

The evolution of mycorrhizae had a profound impact on terrestrial ecosystems. Fossils of arbuscular mycorrhizae from the Rhynie chert (400 million years before present) suggest that mycorrhizae facilitated the colonization of the land by plants, and today over 90% of plants form mycorrhizae of some kind^{6,7}. The majority of mycorrhizae are arbuscular mycorrhizae, which involve the monophyletic Glomales and a broad range of herbaceous and woody plants⁸. The other major group of mycorrhizae are ectomycorrhizae, which are formed primarily by homobasidiomycetes, as well as a handful of ascomycetes and zygomycetes⁸. About thirty families of plants form ectomycorrhizae, including pines, oaks, dipterocarps and eucalypts⁸. Forests of these and other ectomycorrhizal trees dominate many cool temperate and some tropical ecosystems⁷.

We inferred the evolutionary history of ectomycorrhizae among the homobasidiomycetes using a dataset of 161 species that includes 46 ectomycorrhizal species (29%), which is comparable to the proportion of known ectomycorrhizal taxa (about 35%)⁹. Char-

acters for phylogenetic analyses¹⁰ were obtained from overlapping datasets of nuclear small subunit (nuc-ssu) ribosomal DNA (rDNA), nuclear large subunit (nuc-lsu) rDNA and mitochondrial small subunit (mt-ssu) rDNA. Phylogenetic analysis yielded over 10,000 equally parsimonious trees (10,014 steps, consistency index, CI = 0.260; Fig. 1). Despite the large number of equally parsimonious trees, the strict consensus tree is highly resolved (Fig. 1).

Parsimony-based ancestral state reconstructions¹¹ on all equally parsimonious trees imply that there have been 15–16 transformations between ectomycorrhizal and free-living forms, including 7–13 gains and 3–9 losses (average: 9.0 gains, 6.3 losses). The ambiguity in the precise number of gains and losses is due to topological differences among the equally parsimonious trees, as well as the existence of several equally parsimonious reconstructions of ancestral states on all trees. Nevertheless, all reconstructions agree that the ancestor of the homobasidiomycetes was free-living, and that there have been many gains and losses of ectomycorrhizal symbioses within the homobasidiomycetes. This conclusion is also supported by maximum likelihood estimates of ancestral states¹² at 15 key nodes (Fig. 1). (Additional analyses are described in the Supplementary Information.)

Ectomycorrhizal homobasidiomycetes occur in six independent clades, which we refer to as groups 1–6 (Fig. 1). The plants that form ectomycorrhizae also make up a polyphyletic group, including conifers and multiple clades of angiosperms^{8,13}. To evaluate broad patterns of host specificity, we examined the phylogenetic distribution of associations with six of the most ecologically important groups of ectomycorrhizal plants: Betulaceae, Dipterocarpaceae, Fagaceae, Myrtaceae, Pinaceae and Salicaceae. All six groups of ectomycorrhizal homobasidiomycetes form associations with Pinaceae, and four groups are reported to form associations with at least four of the angiosperm families that we investigated^{14–19} (Fig. 1). At low taxonomic levels, there is considerable variation in the degree of host specificity exhibited by ectomycorrhizal fungi and plants, with both groups including generalist and specialist species^{14,20}. It appears, however, that the majority of ectomycorrhizal fungi and plants are capable of entering into symbioses with multiple partners. For example, western hemlock, *Tsuga heterophylla*, forms associations with over 100 fungal species²¹, which represent five of the six groups of ectomycorrhizal fungi that we recognize here, and the fly agaric, *Amanita muscaria*, forms associations with at least 23 species of Betulaceae, Fagaceae, Pinaceae and Salicaceae¹⁴. This broad host potential contributes to a dynamic community structure, in which individual plants form associations with suites of fungal species, whose composition may shift over time^{7,8,22}.

Adding to the dynamism of ectomycorrhizal assemblages is the occasional occurrence of reversals to a free-living, saprotrophic lifestyle within otherwise ectomycorrhizal clades of fungi. Our results suggest that as many as nine such reversals have occurred, although the precise number is not resolved with parsimony (Fig. 1). One unambiguous reversal occurs on the lineage leading to *Lentaria*, which is a small group of about 15 species of wood-decaying fungi that is nested in group 5 (Fig. 1). However, by far the largest concentration of secondarily saprotrophic taxa is in group 1, which is composed of three clades that have been termed the euagaric, bolete and theleporoid clades²³ (Fig. 1). Collectively, it has been estimated that these clades include about 9,500 described species²³, which is 73% of the roughly 13,000 known species of homobasidiomycetes⁹. On the basis of current taxonomy, we estimate that there are about 8,500 described species of saprotrophic homobasidiomycetes, and that 4,700 of these species are nested in group 1. Thus, about half of all saprotrophic homobasidiomycetes may ultimately have been derived from ectomycorrhizal ancestors.

Saprotrophic homobasidiomycetes include two important (somewhat overlapping) guilds: wood-decay fungi, and soil and leaf-litter fungi²⁴. Both play pivotal roles in nutrient cycling in terrestrial ecosystems. Ninety-four species in our data set, including

† Present address: Department of Biology, Clark University, 950 Main Street, Worcester, Massachusetts 01610, USA (D.S.H.); Laboratory of Genetics, Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, California 92037-1099, USA (L.-B. G.); Department of Ecology and Evolutionary Biology, Yale University, P.O. Box 208106, New Haven, Connecticut 06520-8106, USA (M.J.D.).

both outgroup species, decay wood, which suggests that wood decay is the plesiomorphic condition in the homobasidiomycetes. Of the fourteen litter-decaying fungi in our data set, all but two are secondary saprotrophs in the euagarics clade (Fig. 1). In both ectomycorrhizal and litter-decaying fungi, the vegetative mycelia

proliferate in the upper layers of soil, and fruiting bodies are produced directly on soil²⁵. Therefore, derivation of litter decayers from ectomycorrhizal forms would require only a switch from a symbiotic to a free-living lifestyle, without a change of habitat.

Our inference that multiple lineages of homobasidiomycetes have 'escaped' from ectomycorrhizal symbioses is consistent with theoretical predictions that mutualisms are inherently unstable^{2,3}. This view is also supported by empirical studies, which have demonstrated that symbioses can shift along a continuum of interactions ranging from mutualism to antagonism^{26,27}. The situation in homobasidiomycetes is unusual, however, in that the transitions have not been between mutualists and parasites, but between mutualists and free-living forms⁴. Significantly, the eight timber pathogens in our data set do not appear to be closely related to ectomycorrhizal taxa (Fig. 1). The derivation of free-living forms from mutualists has received little attention, but might be common in groups where symbionts have some capacity for living independently and hosts have multiple potential symbionts. Besides ectomycorrhizae, such associations might include those involving corals and their photosynthetic zooxanthellae, or leaf-cutter ants and their cultivated fungi.

Several factors are thought to promote the stability of mutualisms, including compensating mechanisms that limit exploitation of one partner by another⁵; strictly vertical transmission of symbionts between host generations; asexual reproduction of symbionts; and one-to-one correspondence of host and symbiont species^{3,4}. Based on these criteria, the apparent instability of ectomycorrhizae is not surprising. There are no known mechanisms that protect the fungi from exploitation by their plant hosts; fungal symbionts reproduce sexually and disperse spores independently of plant propagules; and most ectomycorrhizal plants draw partners from a diverse pool of potential symbionts. One factor that might be expected to promote the stability of ectomycorrhizal associations is the putative dependence of the fungus on the host for photosynthate. The view that ectomycorrhizal fungi are obligate symbionts is based on the slow growth of many of the fungi in pure culture, and their inability to derive adequate nutrition for growth from complex organic substrates (for example, cellulose)³⁵. However, cultural studies have shown that certain ectomycorrhizal fungi can degrade cellulose, hemicellulose, pectin and lignin, albeit to a limited extent, and may act saprotrophically in the soil^{25,28}. It appears that some ectomycorrhizal fungi have retained the genes for enzymes that can degrade plant cell walls, which may have facilitated repeated evolutionary reversals to saprotrophy.

Methods

Sequence data were obtained by direct automated cycle sequencing of polymerase chain reaction products. Sequences generated for this study include 13 nuc-ssu rDNA, 10 mt-ssu rDNA and 50 nuc-lsu rDNA sequences, which were aligned with published sequences using Clustal X and adjusted by eye. Hypervariable regions that were deemed too divergent to align were excluded from analyses. The dataset included 157 nuc-ssu sequences, 158 mt-ssu sequences, and 60 nuc-lsu sequences, with 3,251 aligned positions, of which 1,111 are parsimony informative. Phylogenetic analyses using equally weighted parsimony were conducted using PAUP* 4.0 (ref. 11), with 1,000 heuristic searches with random taxon addition sequences, the maximum number of trees (MAXTREES) limited to 10,000,

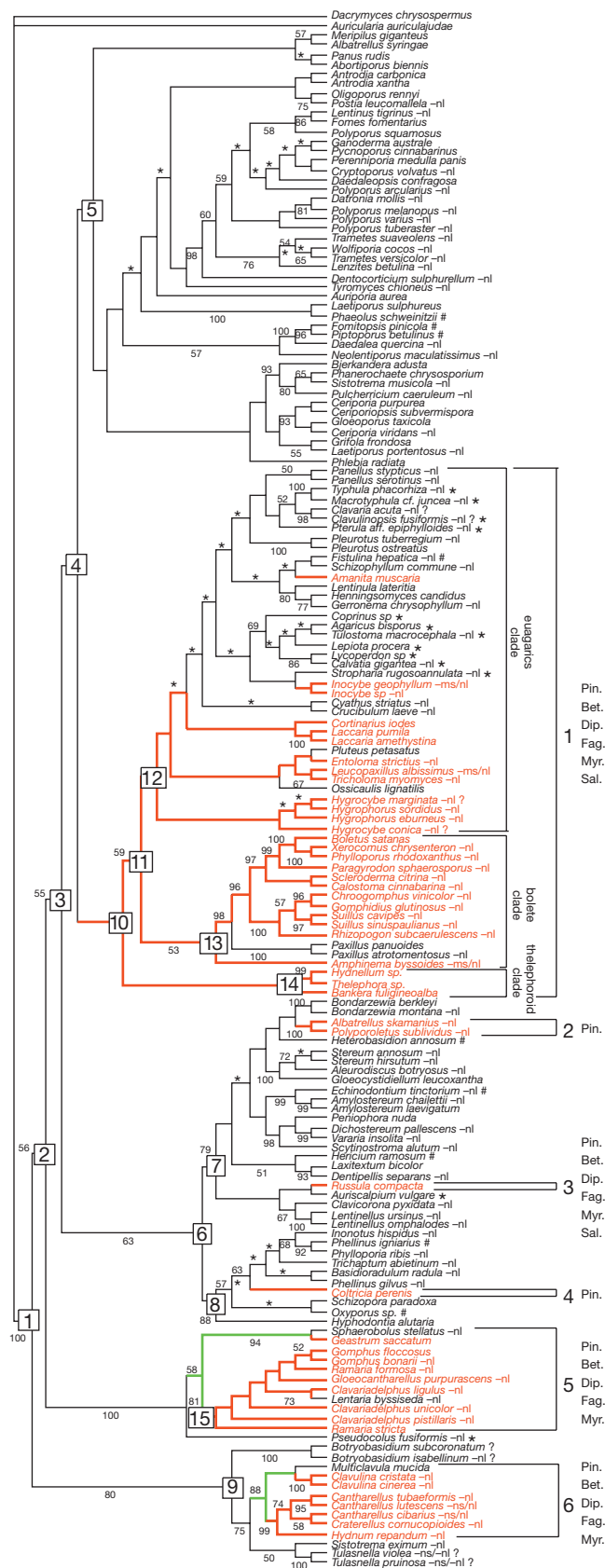


Figure 1 Phylogeny of homobasidiomycetes and evolution of ectomycorrhizal symbioses inferred from rDNA sequences. One of 10,000 equally parsimonious trees. Branches that collapse in the strict consensus tree are marked with asterisks. Bootstrap frequencies over 50% are shown. Ancestral states according to parsimony are indicated with branch shading: red is ectomycorrhizal, black is non-ectomycorrhizal, green is uncertain. Using maximum likelihood, nodes 1–9 are estimated to be non-ectomycorrhizal, and nodes 10–15 are estimated to be ectomycorrhizal ($\Delta\log L > 2$). Species are marked with symbols: coded as uncertain in parsimony optimization (questionmarks); litter-decaying saprotrophs (asterisks); tree pathogens (hashes); –ms means no mt-ssu rDNA sequence; –ns means no nuc-ssu rDNA sequence; –nl means no nuc-lsu rDNA sequence. Host plant codes: Pin., Pinaceae; Bet., Betulaceae; Dip., Dipterocarpaceae; Fag., Fagaceae; Myr., Myrtaceae; Sal., Salicaceae.

and tree bisection-reconnection (TBR) branch swapping. Bootstrap analyses used 100 heuristic searches, with one random taxon addition sequence per replicate, and with MAXTREES limited to 10. Based on previous analyses²⁹, the heterobasidiomycetes *Auricularia* and *Dacrymyces* were used for rooting purposes. Terminal taxa were coded as ectomycorrhizal, non-ectomycorrhizal or uncertain, and ancestral state reconstructions were performed on all trees with equally weighted parsimony using MacClade 3.0 (ref. 11). Maximum likelihood estimates of support for alternative states at selected nodes of one tree selected at random (with branch lengths estimated from molecular data using parsimony) were calculated using Discrete¹². Because Discrete requires that all terminal taxa be scored for the character of interest, the estimated states from parsimony optimizations were assigned for the eight species that were coded as uncertain (Fig. 1). The 'local' method for estimating support for alternative ancestral states was used, with a difference of two units in log likelihood ($\Delta\log L > 2$) taken as an approximate criterion of significance¹².

Received 14 March; accepted 20 July 2000.

- Price, P. W. In *Symbiosis As A Source Of Evolutionary Adaptation* (eds Margulis, L. & Fester, R.) 262–272 (MIT Press, Cambridge, MA, 1991).
- Bronstein, J. L. Conditional outcomes in mutualistic interactions. *Trends Ecol. Evol.* **9**, 214–217 (1994).
- Thompson, J. N. *The Coevolutionary Process* (Univ. Chicago Press, Chicago, 1994).
- Herre, E. A., Knowlton, N., Mueller, U. G. & Rehner, S. A. The evolution of mutualisms: exploring the paths between conflict and cooperation. *Trends Ecol. Evol.* **14**, 49–53 (1999).
- Pellmyr, O. & Huth, C. J. Evolutionary stability of mutualism between yuccas and yucca moths. *Nature* **372**, 257–260 (1994).
- Pirozynski, K. A. & Malloch, D. W. The origin of land plants: a matter of mycotrophism. *Biosystems* **6**, 153–164 (1975).
- Malloch, D. W., Pirozynski, K. A. & Raven, P. H. Ecological and evolutionary significance of mycorrhizal symbioses in vascular plants (a review). *Proc. Natl Acad. Sci. USA* **77**, 2113–2118 (1980).
- Smith, S. E. & Read, D. J. *Mycorrhizal Symbiosis* 2nd edn (Academic, San Diego, 1997).
- Hawksworth, D. L., Kirk, P. M., Sutton, B. C. & Pegler, D. N. *Dictionary Of The Fungi* 8th edn (CAB International, Wallingford, 1996).
- Swofford, D. L. *PAUP* 4.0b2a* (Sinauer, Sunderland, 1999).
- Maddison, W. P. & Maddison, D. R. *MacClade version 3* (Sinauer, Sunderland, 1992).
- Pagel, M. The maximum likelihood approach to reconstructing ancestral character states of discrete characters on phylogenies. *Syst. Biol.* **48**, 612–622 (1999).
- Fitter, A. H. & Moyersoen, B. Evolutionary trends in root-microbe symbioses. *Philos. Trans. R. Soc. (Ser. 351)* **1345**, 1367–1375 (1996).
- Trappe, J. M. Fungus associates of ectotrophic mycorrhizae. *Bot. Rev.* **28**, 538–606 (1962).
- Ashton, D. H. Studies on the mycorrhizae of *Eucalyptus regnans* F. Muell. *Aust. J. Bot.* **24**, 723–741 (1976).
- Malajczuk, N., Molina, R. & Trappe, J. M. Ectomycorrhiza formation in *Eucalyptus*. I. Pure culture syntheses, host specificity and mycorrhizal compatibility with *Pinus radiata*. *New Phytol.* **91**, 467–482 (1982).
- Smits, W. T. M. *Dipterocarpaceae: Mycorrhizae and regeneration* (Tropenbos Foundation, Wageningen, 1994).
- Nuhamara, S. T., Hadi, S. & Baimatmadja, E. I. In *Proceedings of the 6th North American Conference on Mycorrhizae* (ed. Molina, R.) 439 (Forest Research Laboratory, Oregon State Univ., 1985).
- Danielson, R. M. Ectomycorrhizal associations in jack pine stands in north-eastern Alberta. *Can. J. Bot.* **62**, 932–939 (1984).
- Molina, R. & Trappe, J. M. Patterns of ectomycorrhizal host specificity and potential among pacific northwest conifers and fungi. *Forest Sci.* **28**, 423–458 (1982).
- Kropp, B. R. & Trappe, J. M. Ectomycorrhizal fungi of *Tsuga heterophylla*. *Mycologia* **74**, 479–488 (1982).
- Bruns, T. D. Thoughts on the processes that maintain local species diversity of ectomycorrhizal fungi. *Plant Soil* **170**, 63–73 (1995).
- Hibbett, D. S. & Thorn, R. G. In *The Mycota Vol. VII Systematics and Evolution* (eds McLaughlin, D. J., McLaughlin, E. G. & Lemke, P. A.) (Springer, in the press).
- Cooke, R. C. & Rayner, A. D. M. *Ecology of Saprotrophic Fungi* (Longman, New York, 1984).
- Leake, J. R. & Read, D. J. In *The Mycota Vol. IV Environmental and Microbial Relationships* (eds Wicklow, D. T. & Söderström, B. E.) 281–301 (Springer, Berlin, 1997).
- Pellmyr, O., Leebens-Mack, J. & Huth, C. J. Non-mutualistic yucca moths and their evolutionary consequences. *Nature* **380**, 155–156 (1996).
- Clay, K. In *Coevolution of Fungi with Plants and Animals* (eds Pirozynski, K. A. & Hawksworth, D. L.) 79–105 (Academic, San Diego, 1988).
- Lamb, R. J. Effect of D-glucose on utilization of single carbon sources by ectomycorrhizal fungi. *Trans. Br. Mycol. Soc.* **63**, 295–306 (1974).
- Swann, E. C. & Taylor, J. W. Higher taxa of basidiomycetes: an 18S rRNA perspective. *Mycologia* **85**, 923–936 (1993).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank M. Pagel for providing a copy of Discrete; USDA, DAOM and other sources for fungal samples; J.-M. Moncalvo and R. Vilgalys for access to unpublished sequences; and J. Bronstein for helpful comments. This work was supported by grants from the NSF.

Correspondence and requests for materials should be addressed to D.S.H. (e-mail: dhibbett@black.clarku.edu). Sequences have been deposited in GenBank (accession numbers AF287817–AF287891), and the data matrix and tree have been deposited in TreeBASE (<http://phylogeny.harvard.edu/treebase>).

The genome sequence of the thermoacidophilic scavenger *Thermoplasma acidophilum*

Andreas Ruepp*, Werner Graml*, Martha-Leticia Santos-Martinez*, Kristin K. Koretke†, Craig Volker†, H. Werner Mewes‡, Dmitrij Frishman‡, Susanne Stocker‡, Andrei N. Lupas† & Wolfgang Baumeister*

* Max-Planck-Institut für Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, Germany

† Bioinformatics, Smith Kline Beecham Pharmaceuticals, Collegeville, Pennsylvania 19426, USA

‡ GSF-Forschungszentrum für Umwelt und Gesundheit, Munich Information Center for Protein Sequences, Am Klopferspitz 18a, D-82152 Martinsried, Germany

Thermoplasma acidophilum is a thermoacidophilic archaeon that thrives at 59 °C and pH 2, which was isolated from self-heating coal refuse piles and solfatara fields^{1,2}. Species of the genus *Thermoplasma* do not possess a rigid cell wall, but are only delimited by a plasma membrane. Many macromolecular assemblies from *Thermoplasma*, primarily proteases and chaperones, have been pivotal in elucidating the structure and function of their more complex eukaryotic homologues^{3,4}. Our interest in protein folding and degradation led us to seek a more complete representation of the proteins involved in these pathways by determining the genome sequence of the organism. Here we have sequenced the 1,564,905-base-pair genome in just 7,855 sequencing reactions by using a new strategy. The 1,509 open reading frames identify *Thermoplasma* as a typical euryarchaeon with a substantial complement of bacteria-related genes; however, evidence indicates that there has been much lateral gene transfer between *Thermoplasma* and *Sulfolobus solfataricus*, a phylogenetically distant crenarchaeon inhabiting the same environment. At least 252 open reading frames, including a complete protein degradation pathway and various transport proteins, resemble *Sulfolobus* proteins most closely.

Two basic approaches have been taken to genome sequencing. The statistical approach ('shotgun sequencing') relies on the determination of a highly redundant set of random genomic DNA sequences, which are assembled in the computer, the gaps remaining to be closed by other methods⁵. This approach rapidly yields 90–98% of the genomic information, but requires an extensive robotic infrastructure. The directed approach relies on a complete mapping of the genome, followed by the sequencing of large overlapping genomic fragments by shotgun sequencing or primer walking⁶. This approach reduces the infrastructure requirements but is slowed down by the need for a genetic map.

One of our aims was to establish a strategy for sequencing microbial genomes in reasonable time without extensive infrastructure. This strategy ('shotgun primer walking') combines features of the statistical and directed methods. After construction of several phagemid libraries and one cosmid library, inserts from randomly chosen clones were sequenced from the ends by primer walking (see Supplementary Information). Sequencing was stopped in regions that were redundant with already determined sequences. In total, 400 phagemids, covering 850 kilobase-pairs (kbp) (54%), and 469 cosmids, covering 1,533 kbp (98%), were partially or fully sequenced. With an average insert length of 40 kbp, the clones of the cosmid library statistically covered the genome 12 times. Because cosmids are susceptible to recombination events, the reverse strand of cosmid DNA was always sequenced with DNA templates from other cosmid clones covering the same region, or

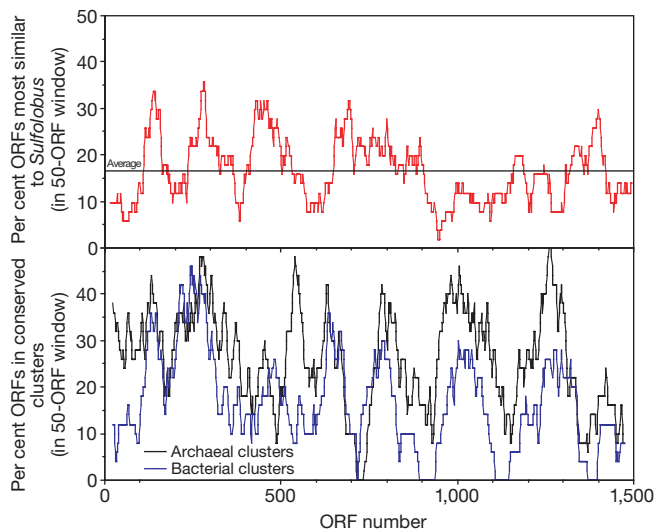


Figure 1 Chromosomal distribution of *Thermoplasma* ORFs most similar to *Sulfolobus* (top panel), and of ORFs found in conserved gene clusters (bottom panel). ORFs were considered to occur in a conserved cluster if their chromosomal vicinity was also observed in at least 1 of 13 complete reference genomes (6 archaeal, 7 bacterial; see Methods). The approximate locations of proteins most similar between *Thermoplasma* and *Sulfolobus* mentioned in the text are indicated in the top panel.

from polymerase chain reaction (PCR) fragments generated from genomic DNA. As the project advanced, fragments were assembled into progressively larger contigs, until 4 gaps of 33 kbp in total were left. These were closed by four PCR fragments, which were sequenced by primer walking. By this method, we completed the 1.56 Mbp genome of *T. acidophilum* with 7,855 sequencing reactions. For comparison, shotgun sequencing of the 1.86 Mbp *Thermotoga maritima* genome, which has the same G+C content, required over 30,000 sequencing reactions.

During the annotation process, we detected ten open reading frames (ORFs) with putative frameshifts. The corresponding genomic regions were amplified by PCR and resequenced with the primers obtained during primer walking. Four of the frameshifts were confirmed, and six had to be corrected. All errors occurred in phagemids, which, unlike cosmids, were sequenced on both strands but without using template DNA from other sources. This suggests that in genome projects operating with low sequence coverage the two strands should be sequenced with template DNA from different sources. It also shows, however, that high-quality data can be obtained with only twofold sequence coverage.

Beyond the advantages discussed here, shotgun primer walking generates an extensive primer set, covering the whole genome. This represents a powerful tool, which can be used to determine rapidly from PCR fragments differences in related DNA sequences, for example between strains of pathogenic bacteria or between individuals in human populations.

The genome of *T. acidophilum* DSM 1728, one of the smallest among free-living organisms, consists of a single circular chromosome of 1,564,905 bp (Table 1). We did not detect any plasmids, either by biochemical methods or through DNA sequencing. A single plasmid of 15.2 kbp has previously been reported in other isolates of *T. acidophilum*⁷. The origin of replication was estimated using the cumulative skew of the orientation of ORFs as well as different nucleotide skews⁸. The centre of a 477-bp intergenic region in this area was chosen as nucleotide 1.

Thermoplasma contains one copy of each of the ribosomal RNAs, which are dispersed in the genome⁹ and separated by at least 52 kbp. This is different from all other known archaea, where at least two of the three genes are contiguous on the chromosome.

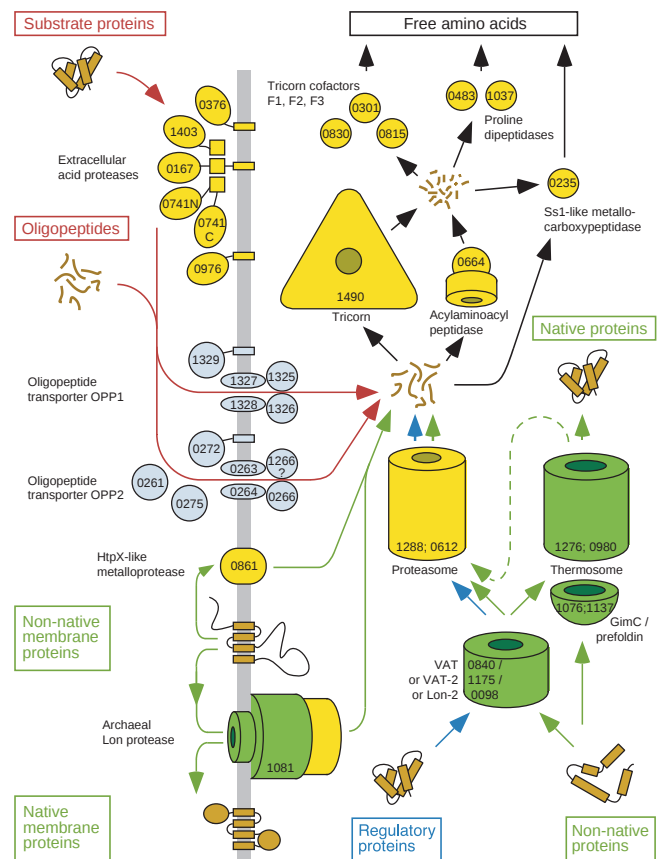


Figure 2 Overview of the main proteolytic pathways in *Thermoplasma*. Red arrows indicate the metabolic pathway, green arrows the 'quality control' pathway for damaged proteins and blue arrows the regulatory pathway. Chaperones are shown in green, proteases in yellow and transporters in blue. The numbers indicate the ORF code; that is, 1490 corresponds to Ta1490.

We identified 1,509 ORFs in the genome (see Supplementary Information), of which over one-third have homologues in all three domains of life (see Supplementary Information). After manual curation, 823 ORFs (55%) were considered similar to functionally annotated proteins, 446 (29%) resembled hypothetical ORFs in other organisms, and 240 (16%) were not recognizably similar to other sequences in current databases ('singletons'). In total, we could assign 685 ORFs (46%) to categories of the MIPS Functional Catalogue (see Supplementary Information). The only category that deviates significantly from other archaea is transport facilitation, which occupies a larger proportion of the genome than in any other archaeon, but is not exceptional when compared with some bacteria (*Thermotoga*, *Deinococcus*). By conservative search methods, we could match 620 domains of 537 ORFs (35.6%) to proteins of known structure (Table 1; Supplementary Information). The distribution of structural classes resembles that of other prokaryotic genomes, but shows a striking overrepresentation of three-layer ($\alpha\beta\alpha$) sandwich folds (43% of all matches), which primarily encompass enzymes.

About one-third of the ORFs (484; 32%) occurred in 139 conserved gene clusters, as judged relative to a set of 7 bacterial and 6 archaeal genomes (see Methods). Twenty-two of these clusters occurred only in *Thermoplasma* and bacteria and are presumably due to lateral transfer. The clusters occur in 'hot spots' on the chromosome (Fig. 1), suggesting that entire gene regions may have been acquired in discrete genetic events. The largest detected clusters encode ribosomal proteins (Ta1249–Ta1271), NADH dehydrogenase subunits (Ta0959–Ta0970), precorrin biosynthesis

proteins (Ta0653–Ta0660) and flagellar proteins (Ta0553–Ta0560).

In the thermoacidophiles *Thermoplasma* and *Sulfolobus*, glucose degradation proceeds by a non-phosphorylated variant of the Entner–Doudoroff pathway¹⁰ (see Supplementary Information), in which the first step is catalysed by glucose dehydrogenase. The acetyl-CoA produced in this pathway enters the oxidative tricarboxylic acid (TCA) cycle. A complete set of TCA enzymes is recognizable in *Thermoplasma*, many of which have been studied experimentally. In the glycolysis/gluconeogenesis pathway (Embden–Meyerhof–Parnas; EMP), homologues are recognizable for most enzymes, but not for phosphofructokinase and fructose

bisphosphate aldolase, so the presence of a working EMP pathway¹¹ cannot be confirmed.

Thermoplasma is microaerophilic and contains several respiratory chain proteins, such as electron transfer flavoproteins (Ta0429, Ta0329 and Ta0212) and cytochrome *b* homologues (Ta1222, Ta1228 and Ta1003). Two of the latter are found in gene clusters with Rieske iron–sulphur proteins (Ta1223 and Ta1229). In addition, we identified two homologues of a cytochrome *bd* quinol oxidase¹² (Ta1484 and Ta0992).

Thermoplasma is also able to gain energy anaerobically by sulphur respiration². Unexpectedly, no homologues of the genes that mediate sulphur respiration in *Archaeoglobus fulgidus*¹³ were found. Instead, *Thermoplasma* contains homologues of AsrA (Ta0046) and AsrB (Ta0047), which mediate dissimilatory sulphur reduction in *Salmonella typhimurium*¹⁴. A sulphide–quinone reductase homologue (Ta1129) may also be involved in sulphur respiration.

Most studied archaea are motile and previous work suggested that they use a chemotaxis signal transduction pathway resembling that of bacteria¹⁵. However, *Thermoplasma* is flagellated and motile², yet lacks recognizable chemotaxis proteins. This situation is also found in other archaea (*Methanococcus jannaschii*, *Aeropyrum pernix*) and bacteria (*Aquifex aeolicus*), suggesting the presence of an alternative, unrelated signal transduction pathway. Another puzzle concerning motility is the lack of a cell wall, which anchors the flagella in prokaryotes. It is unclear what structure could function as the stator for flagellar rotation in *Thermoplasma*.

Two chaperones of *Thermoplasma* have been studied experimentally, the thermosome³ and the protein ‘unfoldase’ VAT^{16,17}. Additional chaperones recognizable in the genome (Table 2) include three Hsp20-like proteins (Ta0437, Ta0471 and Ta0864) and a complete Hsp60 system, consisting of two thermosome subunits (α , Ta0980; and β , Ta1276) and two prefoldin/Gim subunits (α , Ta1076; and β , Ta1137). Unlike many archaea, *Thermoplasma* also contains an Hsp70 system (DnaK, Ta1087; DnaJ, Ta1088; and GrpE, Ta1086), whose genes are clustered on the chromosome in contrast to the endogenous Hsp60 genes.

Thermoplasma contains several ‘unfoldases’ (AAA+ proteins)¹⁷, which are frequently associated with proteolytic degradation¹⁸. These are the Cdc48 homologues VAT (Ta0840) and VAT-2 (Ta1175), an archaeal-type Lon protease (Ta1081) and a Lon-related ATPase lacking the protease domain (Ta0098). Surprisingly, *Thermoplasma* is missing the proteasomal AAA ATPase (PAN; ref. 19), which has hitherto been found in all archaea. It is

Table 1 Genome features

General features of the *Thermoplasma acidophilum* genome

Length of sequence	1,564,905 bp
G+C ratio	46%
Open reading frames	1509
Average ORF length	909 bp
Protein coding regions	87%
Ribosomal RNAs	1 5S; 1 16S; 1 23S
tRNAs	45
7S RNA	1
Repetitive elements	Coordinates
Short, non-coding repeats	846,047–849,104
gtaaaatagaacctaataggattgaaag (29 nt; 46 copies)	
Long repetitive elements	
<i>Ire1</i> (185 nt; 2 copies; 5' flanking sequence of Ta0698 and Ta0708)	739,187–739,371 and 751,166–751,350
<i>Ire2</i> (329 nt; 2 copies; 3' flanking sequence of Ta0698 and Ta0708)	739,451–739,779 and 751,430–751,758
<i>Ire3</i> (654 nt; 2 copies, contains Ta1170/1171 and Ta1414/Ta1415)	1,243,940–1,244,593 and 1,464,457–1,465,110
Long, inverted repeats	
<i>Iir</i> (169 nt; 2 copies)	184,427–184,595 and 397,089–397,257
Chromosomal coding sequences	
No. similar to known proteins	823 (55%)
No. similar to proteins of unknown function	446 (29%)
No. without a database match	240 (16%)
Inteins	1 (Ta004)
No. of protein domains similar to known structures	620 (in 537 proteins)
Mostly α (CATH classification)	19%
Mostly β (CATH classification)	12%
α and β (CATH classification)	69%
No. predicted to contain one transmembrane helix	114 (7.6%)
No. predicted to contain two or more transmembrane helices	268 (17.8%)
No. predicted to contain coiled-coil domains	34 (2%)
nt, nucleotide.	

Table 2 The chaperone complement of *T. acidophilum*

Class	Type	ORF	Annotation	Remarks
Hsp20	Hsp20/sHSP	Ta0437*	ATPase containing hsp20 domain	Hsp20-like C-terminal domain absent in other archaea; ATPase domain similar to <i>E. coli</i> ArsA
		Ta0471	Small heat-shock protein (hsp20) related	
		Ta0864*	Small heat-shock protein (hsp20) related	
Hsp60	Thermosome	Ta0980	Thermosome α -chain	
		Ta1276	Thermosome β -chain	
	GimC	Ta1076	GimC, α -subunit related	
		Ta1137	GimC, β -subunit related	
Hsp70	Hsp70/DnaK	Ta1087*	Probable DnaK-type molecular chaperone	
	Hsp40/DnaJ	Ta1088*	Heat-shock protein DnaJ related	
	Hsp23/GrpE	Ta1086*	Heat-shock protein GrpE related	
AAA+†	Cdc48	Ta0840	VAT ATPase (VCP-like ATPase)	Ta1064 and Ta1217 resemble the double-psi barrel domain of VAT contains only one AAA domain; N-domain does not resemble that of other Cdc48 proteins
		Ta1175	VAT-2 protein	
	Archaeal Lon	Ta1081	ATP-dependent proteinase La (Lon) related	
		Ta0098	Lon-related ATPase	
PDI	Thioredoxin	Ta0866*	Thioredoxin related	Contains two glutaredoxin-like domains
		Ta0125	Glutaredoxin related	
PPlases	FKBP	Ta1011	Peptidyl-prolyl <i>cis-trans</i> isomerase related	
Other	CsaA	Ta1046*	Probable chaperone (csaA)	Similar to the C-terminal domain of archaeal and bacterial Met tRNA synthetases
		Ta1284	Chaperone (csaA) related	

* ORFs of probable bacterial origin.

† *Thermoplasma* contains four other AAA+ proteins with potential chaperone-like activity: Ta0576 and Ta0920, which may act as enzyme-specific factors involved in the maturation of cofactor-containing enzymes; and Ta0799 and Ta1285, which may act in the assembly and/or disassembly of replication complexes.

conceivable that Ta0098, VAT-2 or VAT, which has protein unfolding activity¹⁷, can substitute for this function.

Other chaperones include two protein disulphide isomerases (Ta0125 and Ta0866), a peptidyl-prolyl isomerase of the FKBP family (Ta1011) and two homologues of cold-shock protein CsaA (Ta1284, Ta1046). Ta1284 is most similar to the carboxy-terminal domain of archaeal methionyl-tRNA synthetases (MetRS); this domain is missing from *Thermoplasma* MetRS. Overall, a surprising number of chaperones from *Thermoplasma* appear to have been acquired from bacteria (Table 2).

The pathway for polypeptide degradation in *Thermoplasma* is well understood²⁰. Unfolded proteins are cleaved into fragments of 6–15 residues by the proteasome (α , Ta1288; and β , Ta0612), whose structure and activity have been extensively studied⁴. These fragments are further degraded by the tricorn peptidase (Ta1490) to di-, tri- and tetrapeptides, which are hydrolysed to free amino acids by three tricorn-interacting factors F1 (Ta0830), F2 (Ta0301) and F3 (Ta0815). It is, however, unclear how proteins are targeted to this degradation pathway, because ubiquitin, whose existence in *T. acidophilum* had been proposed on the basis of peptide sequences²¹, is not present in this genome (or indeed in any other archaeon). It is also unclear how proteins are unfolded before degradation by the proteasome, as *Thermoplasma* is missing a proteasomal AAA ATPase.

In addition to the 6 proteins involved in this degradation pathway, *Thermoplasma* contains at least 23 putative proteases (Table 3, Fig. 2), as well as several α/β hydrolases of unknown specificity and a broad-spectrum deacetylase with potential glutamate carboxypeptidase activity (Ta0934). Few proteases seem to be of bacterial origin; however, nine are most similar to proteins of the phylogenetically distant crenarchaeon *Sulfolobus solfataricus* (<http://niji.imb.nrc.ca/sulfhome/results.html>). These proteins are tricorn, its interacting factors and five extracellular, multidomain acid proteases (Ta1403, Ta0167, Ta0376, Ta0741 and Ta0976), three

of which are membrane-anchored by C-terminal transmembrane sequences. It is unclear how *Thermoplasma* retains unanchored secreted proteins without a cell wall, but we note that the two secreted proteases and one of the anchored ones (Ta0167) contain a domain conserved in many archaeal S-layer proteins, which may serve as a scaffold.

Overall, ten *Thermoplasma* proteases are probably extracellular. These include two membrane-anchored signal peptidase I homologues (Ta0151 and Ta0378), which are closely related to their eukaryotic homologues and may yield useful model systems for the more elaborate eukaryotic complex, a membrane-anchored serine peptidase of the S49 family (Ta1083), a probable O-sialoglycoprotein endopeptidase (Ta0324) and a metalloprotease of the M6 family (Ta0728). The two latter proteins are probably secreted, on the basis of experimental findings from other organisms. Like their homologues²², however, they lack recognizable signal sequences. In total, 11 *Thermoplasma* proteases are membrane-associated, including the archaeal Lon protein, which is anchored to the membrane from the cytoplasmic side by a helical hairpin, a structure present in all archaea but missing in bacterial proteins. As the membrane-bound protease FtsH, which is essential in bacteria, is absent from archaea, its function may be assumed by the archaeal Lon protease.

As discussed, *Thermoplasma* surprisingly lacks ubiquitin, a proteasomal ATPase and an archaeal-type sulphur respiration system. We searched systematically for further differences to other archaeal genomes using clusters of orthologous groups (COGs; <http://www.ncbi.nlm.nih.gov/COG>). The first four published archaeal genomes share a 'stable core' of 543 COGs; thirty-six of these appear to be missing in *Thermoplasma* (<http://www.biochem.mpg.de/baumeister/genome>). The most notable are DNA topoisomerase VI and eukaryal-like histones. Topoisomerase VI (ref. 23) is a divergent type II topoisomerase present only in archaea. In its place, *Thermoplasma* contains a bacterial-type DNA gyrase (GyrA, Ta1054; and GyrB, Ta1055).

Table 3 The protease complement of *T. acidophilum*

Clan	Family	ORF	Annotation	Cellular location	Remarks
AX	A5	Ta1403*	Thermopsin related	ext	
		Ta0167*	Thermopsin related	ext/anc	Anchored by C-terminal transmembrane helix
		Ta0741*	Acid protease	ext	Protease domain 1, residues 78–386
	A7	Ta0376*	Pseudomonapepsin related	ext/anc	Anchored by C-terminal transmembrane helix
		Ta0741*	Acid protease	ext	Protease domain 2, residues 810–1,257; family A7 probably belongs to clan SB
		Ta0976*	Pseudomonapepsin related	ext/anc	Anchored by C-terminal transmembrane helix
MA	M1	Ta0301*	Tricorn cofactor F2	cyt	F2 and F3 nearest sequence neighbours
		Ta0815*	Tricorn cofactor F3	cyt	
	M6	Ta0728	Metalloprotease	ext?	Divergent family member, but active-site signature most similar to M6
		Ta0861	Heat-shock protein (HtpX) related	mem	
MG	M24A	Ta1439	Methionine aminopeptidase I related	cyt	
		Ta1037	Proline dipeptidase related	cyt	
	M24B	Ta0483	Proline dipeptidase related	cyt	
		Ta0235†	Carboxypeptidase Ss1 related	cyt	
MH	M40	Ta0324	O-sialoglycoprotein endopeptidase-related	ext?	Contains a C-terminal protein kinase domain
MK	M22	Ta1274	S2P protease-related	mem	Families M50 and M51 probably belong to clan MA
MX	M50	Ta1344	SpoIVFB metalloprotease-related	mem	
		Ta0562	SpoIVFB metalloprotease-related	mem	
SC	S9C	Ta0664	Acylaminoacyl-peptidase related	cyt	
		S33	Prolyl iminopeptidase/Tricorn cofactor F1	cyt	
		Ta0830*	Signal peptidase I related	ext/anc	Ta0151 and Ta0378 nearest sequence neighbours; anchored by N- and C-terminal transmembrane helices
SF	S26B	Ta0151	Signal peptidase I related	ext/anc	
SK	S49	Ta0378	Signal peptidase I related	ext/anc	
		Ta0074	Protease IV related	cyt	
		Ta1083	Serine protease	ext/mem	Contains C-terminal integral membrane domain
SX	S16	Ta1081	ATP-dependent proteinase La (Lon) related	cyt/anc	Transmembrane helical hairpin inserted in the ATPase domain
		S41	Tricorn core protease	cyt	
		Ta1490*	Tricorn core protease	cyt	
PB	T1A	Ta1288	Proteasome α -subunit	cyt	Non-peptidase homologue of the β -subunit
		Ta0612	Proteasome β -subunit	cyt	
	T3	Ta0994†	Glutamyltransferase related	cyt	
UX	U46	Ta0465	Pfpl endopeptidase related	cyt	Probably belongs to cysteine protease family C26

Proteases classified according to the MEROPS classification²² (<http://www.merops.co.uk>). anc, anchored; cyt, cytoplasmic; ext, extracellular; mem, membrane.

* ORFs most similar to proteins of *Sulfolobus solfataricus*.

† ORFs of probably bacterial origin.

Thermoplasma also contains proteins not present in any other archaeal genome (68, after elimination of singletons; <http://www.biochem.mpg.de/baumeister/genome>). Most noteworthy was Hta (Ta0093), the first archaeal DNA-binding protein to be investigated²⁴, which is closely related to bacterial proteins and appears to substitute functionally for the missing histones. Other proteins include an *egghead* homologue (Ta0213), a membrane-anchored protein kinase (Ta0488) and a Ras-like GTPase (Ta1192).

In the past, a *Thermoplasma*-like organism has been debated as a possible ancestor of the eukaryotic cytoplasm. The genome sequence, however, shows that *Thermoplasma* is a typical archaeon with a fairly large protein complement of bacterial origin. We could not identify any of the proteins peculiar to eukaryotes, such as subunits of the nuclear pore complex, the exocyst or the cytoskeleton. However, *Thermoplasma* has been reported to contain a cytoskeleton²⁵, and we identified in genome searches a cytosolic coiled-coil protein (Ta1488), which is conserved in all archaea and seems analogous to intermediate filament proteins.

During the comparison to other archaea we noted that 252 *Thermoplasma* ORFs (17%) were most similar to proteins of *S. solfataricus*, a higher number than for any other organism in our reference set—even though this genome sequence is still incomplete. Of these, 60 appear to function in transport and are frequently clustered (for example, Ta1325–Ta1329, Ta0261–Ta0275 and Ta0126–Ta0146). Many others are involved in metabolism (including energy metabolism) or in proteolysis, or are secreted or membrane-bound ORFs of unknown function. With a few notable exceptions, such as DNA polymerase (Ta0450) or Rad50 (Ta0157), hardly any are involved in information processes (replication, transcription and translation). Although it has been observed previously that in prokaryotes core information processing generally tracks organism phylogeny, whereas metabolism is strongly affected by lateral transfer, it is highly unusual to see such a large number of genes transferred between two phylogenetically distant organisms. We propose that the adaptation to an extreme environment shared by few other organisms has led to substantial genetic exchange. This may have proceeded by only a few large genetic events, resulting in the skewed distribution of *Sulfolobus*-like genes in the *Thermoplasma* genome (Fig. 1).

Thermoplasma inhabits a hot and highly acidic environment, sometimes as low as pH 0.5, in which few organisms are viable. It has adapted to scavenging nutrients from the decomposition of organisms killed by the extreme acidity and requires yeast, bacterial or meat extract when grown in culture. There is an absolute growth requirement for basic oligopeptides²⁶ and analysis of the genome shows the presence of a complete degradation chain for exogenous polypeptides, starting with a set of large extracellular proteases, over two oligopeptide transport systems, to a cytosolic degradation machine consisting of tricorn and its cofactors (Fig. 2). All proteins of this degradation chain are most similar to cognate proteins of *Sulfolobus*, as are many other proteins involved in transport and metabolism. Thus, it would seem that two classes of genes can be distinguished in *Thermoplasma*: One is primarily composed of constitutive, 'housekeeping' genes, which generally reflect the phylogenetic origin of the organism. The other mostly contains 'life-style' genes, including but not limited to metabolism, which are tailored to a specific environment and are shared between phylogenetically distant organisms within one ecological niche.

Methods

Genome sequencing analysis methods

Analysis methods are given in the Supplementary Information.

Annotation

In a first approximation, gene prediction was performed automatically by the Orpheus software²⁷ allowing for ORFs longer than 150 bases and for overlaps between ORFs no longer than 30% of the length of the shorter overlapping ORF. As an additional precaution

we modified the algorithm to keep all ORFs longer than 450 bases as part of the preliminary ORF set destined for manual inspection. Selection of putative gene starts was assisted by ribosome-binding-site detection; however, the information content of these sites in *T. acidophilum* proved to be insufficient for improving prediction results. The predicted ORF complement was then manually refined on the basis of extrinsic evidence.

The main vehicle for automatic and manual annotation of gene products was the PEDANT software suite²⁸. Complete annotation of the genome, including DNA and protein viewers, extensive protein reports, multiple functional and structural categories and search capabilities is available at <http://pedant.mips.biochem.mpg.de>.

Global comparisons of gene complement, searches for conserved gene clusters, and in-depth annotation of protein families were performed in the SmithKline Beecham AiBi Toolkit, a suite of software tools running on top of a relational database containing 68 partial and complete genomes of archaeal (8), bacterial (55) and eukaryotic origin (5), in addition to the non-redundant public sequence database from NCBI (<http://www.ncbi.nlm.nih.gov>). Sequence comparisons were made using BLAST2 and PSI-BLAST²⁹, with low-complexity regions (including coiled coils and transmembrane regions) masked out and a significance threshold of e^{-3} . Conserved gene clusters were identified by comparing pairs of ORFs separated by at most three ORFs between *Thermoplasma* and a reference set of six archaeal and seven bacterial genomes, chosen to broadly represent archaea and deep-branching bacteria. The archaeal genomes were *Methanococcus jannaschii*, *Methanobacterium thermoautotrophicum*, *Archaeoglobus fulgidus*, *Pyrococcus abyssii*, *Pyrococcus furiosus* and *Aeropyrum pernix*; and the bacterial genomes were *Escherichia coli*, *Bacillus subtilis*, *Clostridium acetobutylicum*, *Thermotoga maritima*, *Aquifex aeolicus*, *Deinococcus radiodurans* and *Synechocystis* sp. For opening a cluster, both members of the ORF pair had to have BLAST2 matches in chromosomal vicinity at e^{-10} in at least one of the reference genomes. Clusters were then extended by relaxing the significance cutoff to e^{-3} for proteins satisfying the distance cutoff. Finally, clusters sharing at least one ORF were merged.

Protein fold predictions were made using the PSI-Blast-based search routine SENSER³⁰, and results were compared against the CATH classification (see Supplementary Information). We report only the PSI-Blast results because, with the chosen settings, they have an estimated error rate of less than 5%. However, SENSER returned predictions for an additional 152 proteins with stringent settings, and 205 proteins with relaxed settings, for a total of 727 proteins (48% of the ORF complement; <http://www.biochem.mpg.de/baumeister/genome>). Coiled coils were predicted with the COILS program (http://www.ch.embnet.org/software/COILS_form.html), and signal sequences with SignalP (<http://www.cbs.dtu.dk/services/SignalP/>) and PHYSEAN.

Received 13 April; accepted 6 July 2000.

- Darland, G., Brock, T. D., Samsonoff, W. & Conti, S. F. A thermophilic acidophilic mycoplasma isolated from a coal refuse pile. *Science* **170**, 1416–1418 (1970).
- Segerer, A. & Stetter, K. O. in *The Prokaryotes* (eds Balows, A., Trüper, H. G., Dworkin, M., Harder, W. & Schleifer, K. H.) 712–718 (Springer, New York, 1992).
- Gutsche, L., Essen, L. O. & Baumeister, W. Group II chaperonins: New Tric(k)s and Turns of a Protein Folding Machine. *J. Mol. Biol.* **293**, 295–312 (1999).
- Voges, D., Zwickl, P. & Baumeister, W. The 26S proteasome: A molecular machine designed for controlled proteolysis. *Annu. Rev. Biochem.* **68**, 1015–1068 (1999).
- Fleischmann, R. D. et al. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**, 496–512 (1995).
- Himmelreich, R. et al. Complete sequence analysis of the genome of the bacterium *Mycoplasm pneumoniae*. *Nucleic Acids Res.* **24**, 4420–4449 (1996).
- Yasuda, M., Yamagishi, A. & Oshima, T. The plasmids found in isolates of the acidothermophilic archaeobacterium *Thermoplasma acidophilum*. *FEMS Microbiol. Lett.* **128**, 157–161 (1995).
- Lopez, P., Philippe, H., Myllykallio, H. & Forterre, P. Identification of putative chromosomal origins of replication in Archaea. *Mol. Microbiol.* **32**, 881–891 (1999).
- Ree, H. K. R. & Zimmermann, R. A. Organization and expression of the 16S, 23S and 5S ribosomal RNA genes from the archaeobacterium *Thermoplasma acidophilum*. *Nucleic Acids Res.* **18**, 4471–4478 (1990).
- Budgen, N. & Danson, M. J. Metabolism of glucose via a modified Entner–Doudoroff pathway in the thermoacidophilic archaeobacterium *Thermoplasma acidophilum*. *FEBS Lett.* **196**, 207–210 (1986).
- Searcy, D. G. & Whalley, F. R. *Thermoplasma acidophilum*: Glucose degradative pathways and respiratory activities. *Syst. Appl. Microbiol.* **5**, 30–40 (1984).
- Luebben, M. Cytochromes of archaeal electron transfer chains. *Biochim. Biophys. Acta* **1229**, 1–22 (1995).
- Klenk, H. P. et al. The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. *Nature* **390**, 364–370 (1997).
- Huang, C. J. & Barrett, E. L. Sequence Analysis and Expression of the *Salmonella typhimurium* *asr* Operon Encoding Production of Hydrogen Sulfide from Sulfite. *J. Bacteriol.* **173**, 1544–1553 (1991).
- Rudolph, J. & Oesterhelt, D. Chemotaxis and phototaxis require a CheA histidine kinase in the archaeon *Halobacterium salinarum*. *EMBO J.* **14**, 667–673 (1995).
- Coles, M. et al. The solution structure of VAT-N reveals a 'missing link' in the evolution of complex enzymes from a simple beta alpha beta beta element. *Curr. Biol.* **9**, 1158–1168 (1999).
- Golbik, R., Lupas, A. N., Koretke, K. K., Baumeister, W. & Peters, J. The janus face of the archaeal Cdc48/p97 homologue VAT: Protein folding versus unfolding. *Biol. Chem.* **380**, 1049–1062 (1999).
- Lupas, A., Flanagan, J. M., Tamura, T. & Baumeister, W. Self-compartmentalizing proteases. *Trends Biochem. Sci.* **22**, 399–404 (1997).
- Zwickl, P., Ng, D., Woo, K. M., Klenk, H. P. & Goldberg, A. L. An archaeobacterial ATPase, homologous to ATPases in the eukaryotic 26 S proteasome, activates protein breakdown by 20 S proteasomes. *J. Biol. Chem.* **274**, 26008–26014 (1999).
- Tamura, N., Lottspeich, F., Baumeister, W. & Tamura, T. The role of Tricorn protease and its aminopeptidase-interacting factors in cellular protein degradation. *Cell* **95**, 637–648 (1998).
- Wolf, S., Lottspeich, F. & Baumeister, W. Ubiquitin found in the archaeobacterium *Thermoplasma acidophilum*. *FEBS Lett.* **326**, 42–44 (1993).

22. Barrett, A. J., Rawlings, N. D. & Woessner, J. F. *Handbook of proteolytic enzymes* (Academic, San Diego, CA, 1999).
23. Bergerat, A. et al. An atypical topoisomerase II from archaea with implications for meiotic recombination. *Nature* **386**, 414–417 (1997).
24. Stein, D. B. & Searcy, D. G. Physiologically important stabilization of DNA by a prokaryotic histone-like protein. *Science* **202**, 219–221 (1978).
25. Hixon, W. G. & Searcy, D. G. Cytoskeleton in the archaeobacterium *Thermoplasma acidophilum*? Viscosity increase in soluble extracts. *BioSystems* **29**, 151–160 (1993).
26. Smith, P. F., Langworthy, T. A. & Smith, M. R. Polypeptide nature of growth requirement in yeast extract for *Thermoplasma acidophilum*. *J. Bacteriol.* **124**, 884–892 (1975).
27. Frishman, D., Mironov, A., Mewes, H. W. & Gelfand, M. Combining diverse evidence for gene recognition in completely sequenced bacterial genomes. *Nucleic Acids Res.* **26**, 2941–2947 (1998).
28. Frishman, D. & Mewes, H. W. PEDANTIC genome analysis. *Trends Genet.* **13**, 415–416 (1997).
29. Altschul, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
30. Koretke, K. K., Russell, R. B., Copley, R. R. & Lupas, A. N. Fold recognition using sequence and secondary structure information. *Proteins Struct. Funct. Genet.* **37**, 141–148 (1999).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank P. Forterre and P. Lopez for helping to define the origin of replication; M. Boicu and C. Czoppelt for sequencing; G. Mannhaupt for annotating part of the ORFs; I. Echabe for preparing template DNA and sequencing; and B. Marshall for developing software for gene cluster analysis and for data management.

Correspondence and requests for materials should be addressed to W.B. (e-mail: baumeist@biochem.mpg.de). The *Thermoplasma acidophilum* genome sequence has been deposited in the EMBL database (accession number AL139299).

An SNP map of the human genome generated by reduced representation shotgun sequencing

David Altshuler^{*†}, Victor J. Pollara^{*}, Chris R. Cowles^{*}, William J. Van Etten^{*}, Jennifer Baldwin^{*}, Lauren Linton^{*} & Eric S. Lander^{*‡}

^{*} Whitehead Institute/MIT Center for Genome Research, Nine Cambridge Center, Cambridge, Massachusetts 02142, USA

[†] Diabetes Unit, Department of Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts 02114, USA

[‡] Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

Most genomic variation is attributable to single nucleotide polymorphisms (SNPs), which therefore offer the highest resolution for tracking disease genes and population history^{1–3}. It has been proposed that a dense map of 30,000–500,000 SNPs can be used to scan the human genome for haplotypes associated with common diseases^{4–6}. Here we describe a simple but powerful method, called reduced representation shotgun (RRS) sequencing, for creating SNP maps. RRS re-samples specific subsets of the genome from several individuals, and compares the resulting sequences using a highly accurate SNP detection algorithm. The method can be extended by alignment to available genome sequence, increasing the yield of SNPs and providing map positions. These methods are being used by The SNP Consortium, an international collaboration of academic centres, pharmaceutical companies and a private foundation, to discover and release at least 300,000 human SNPs. We have discovered 47,172 human SNPs by RRS, and in total the Consortium has identified 148,459 SNPs. More broadly, RRS facilitates the rapid, inexpensive construction of SNP maps in biomedically and agriculturally important species. SNPs discovered by RRS also offer unique advantages for large-scale genotyping.

To discover SNPs, several copies of each locus must be sampled from a population and compared for sequence differences. Two methods have been described. The first, locus-specific polymerase chain reaction (PCR) amplification (LSA), requires the synthesis of oligonucleotide primers for each locus, limiting it to regions of known sequence and making it expensive for large-scale approaches. Moreover, LSA produces diploid genotypes; this requires identification of SNPs as heterozygotes, which is technically challenging. The second, whole-genome shotgun⁷, sequences random clones from the genomes of many individuals. It does not require previous knowledge of genomic sequence nor PCR, and provides haploid genotypes. Whole-genome shotgun is inflexible, however, requiring several-fold coverage of the genome before SNPs are discovered. For greater flexibility and efficiency, we sought to use 'reduced representations'—reproducibly prepared subsets of the genome, each containing a manageable number of loci to facilitate re-sampling. This approach presented three key challenges: creating reduced representations, finding orthologous matches and obtaining sufficient accuracy for automated calling of SNPs.

Many properties could be used to prepare reduced representations (for example, binding to a particular protein or functioning as an origin of replication). One of the simplest is to purify restriction fragments in a given size range. For example, *Bgl*II sites occur on average every ~3,100 base pairs (bp) in human DNA, which means that restriction fragments with length between 500 and 600 bp should number ~26,000 and comprise ~0.5% of the human genome. Computer analysis of 517 megabases (Mb) of finished human genomic sequence (17% of the estimated 3.1 gigabases) yielded 3,847 *Bgl*II fragments in this range—within 10% of the expected value. Thus, SNPs could be discovered by mixing DNA from many individuals, preparing a library of appropriately sized restriction fragments, and randomly sequencing clones. In this example, 52,000 sequences would provide, on average, twofold coverage of each locus, yielding thousands of SNPs.

We developed rules to align only those sequences representing the same genomic locus, excluding spurious matches arising from repeats. These rules eliminated known repeats, partial alignments that failed to extend across the entire sequence, matches showing excessive sequence divergence (compared with orthologous loci), and sequences re-sampled more often than expected for a single-copy locus (see Methods). Re-sequencing experiments confirmed that these rules successfully eliminated most spurious matches owing to paralogous repeats.

In comparing single-pass sequences for SNPs, base-calling errors can dominate the low rate of true polymorphism. With true SNPs occurring at a rate of 1 in 1,300 bp, base-calling errors must be less than 1 in 52,000 to achieve less than 5% false positive SNPs. Computer programs that estimate sequence accuracy (such as PHRED^{8,9}) judge only a small fraction of single-pass bases to be this accurate. However, we noted that many base-calling 'errors' occur adjacent to easily detected artefacts (that is, compressions and stops), or are attributable to poor alignment. We hypothesized that bases surrounded by perfectly aligned, consistently high-quality sequence (termed 'good neighbourhoods') might be more accurate than predicted by PHRED. We defined a neighbourhood quality standard (NQS) to identify such bases (see Methods).

To test the NQS, we examined 3.8 Mb of single-pass human DNA sequence obtained from bacterial artificial chromosomes (BACs), and compared the base calls to the highly accurate finished sequence of each BAC. Because BAC DNA is clonal, there are no polymorphisms, and any 'SNPs' represent base-calling errors. As expected, PHRED quality scores accurately reflect the overall likelihood of error (see below). Critically, base calls within 'good neighbourhoods' were much more accurate than predicted by PHRED: 85% of bases with PHRED scores more than 20 satisfied the NQS, and displayed an error rate of 1 in 36,000. In contrast, 15% of such bases failed the NQS, and these had a 40-fold higher error rate of 1 in

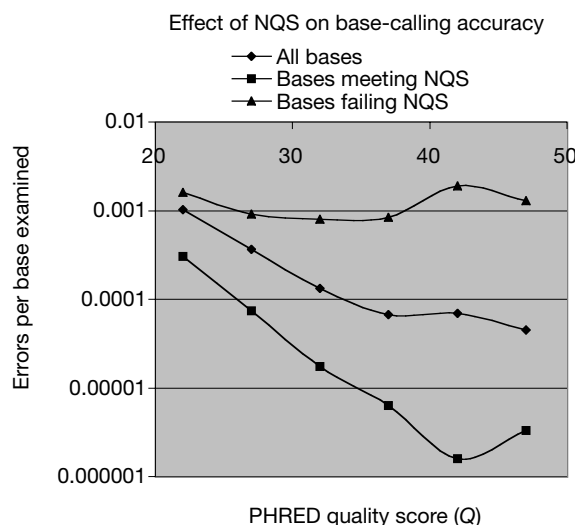


Figure 1 Impact of quality criteria on error rates. Data are plotted according to the PHRED *Q* score of each base^{8,9} and reported in bins of five PHRED *Q* units; only base substitution errors were counted. As previously reported⁹, overall PHRED scores accurately predict the observed rates of base-calling errors; however, bases meeting the NQS display

substantially lower error rates than are predicted by their PHRED scores. Although the effect is proportionally greatest for high PHRED scores, the bulk of errors avoided are found in bases with lower PHRED scores (that is, those with the highest error rates).

825 bp. Improved accuracy was observed across all PHRED scores (Fig. 1). The NQS can be used for automated SNP identification from single-pass sequence from RRS, genomic alignment, or expressed sequence tags^{10–12}. Its accuracy exceeds current goals for finished human sequence (less than 1 error in 10,000 bases) without several-fold coverage of each base; thus, the approach can potentially improve the efficiency and accuracy of genomic sequencing.

As a pilot, we pooled genomic DNA from ten individuals, digested it with *Bgl*II, separated the fragments by size by using agarose gel electrophoresis, excised a narrow slice adjacent to 564 bp, and cloned the fragments. Initial sequencing confirmed a tight distribution of insert lengths (Fig. 2a), and we sequenced 18,720 clones. After applying the alignment and base-calling rules above, 10,588 sequences were matched to one or more other clones, generating 14,400 paired reads that clustered into 3,356 'cliques'. The effective complexity of this library (the reciprocal of the chance that two randomly chosen clones derive from the same locus) is roughly 12,000 unique inserts, consistent with expectations. The distributions of clique sizes and SNPs within cliques were similar to those predicted under composite Poisson sampling, assuming a constant rate of nucleotide diversity, neutral alleles and a constant-sized population (Fig. 2b–d). These distributions, however, had longer than expected tails (Fig. 2b, d), which probably reflects cliques derived from repeats. Such cliques were withheld from further analysis; whether they represent unique loci will be determined by alignment to the complete human genome sequence. Sequence comparison of remaining cliques identified 1,750 candidate SNPs, with a heterozygosity (π) of 7.2×10^{-4} (consistent with previous studies^{13–16}).

Although RRS does not require prior knowledge of genome sequence, it can be extended by genomic alignment (GA). When RRS data were compared with 517 Mb (~16%) of finished human genomic sequence, 16% of clones matched a genomic contig, and SNPs were discovered at an equivalent rate ($\pi = 7.2 \times 10^{-4}$). Draft-quality genomic data can be used by applying the NQS to the submitted PHRAP quality scores (data not shown). Moreover, each SNP discovered by GA is immediately mapped to its location in the genome. Thus, the efficiency of SNP discovery and mapping is substantially increased by the availability of genomic sequence.

On the basis of these and related efforts at The Sanger Centre¹⁷ and Washington University in St. Louis (J. McPherson, personal

communication), The SNP Consortium (TSC) adopted RRS to identify 300,000 human SNPs over 2 years. TSC projects use the anonymous NIH diversity panel of 24 individuals¹⁸, from which we constructed 20 libraries using 4 different enzymes (*Bgl*II, *Hind*III, *Xba*I and *Eco*RI). We had identified 47,172 SNPs (Table 1) at the time of this writing, with the overall total of the TSC being 148,459, which shows that RRS is robust and scaleable. Each SNP is mapped by comparison with public human genome sequence (L. Stein, personal communication). Consistent with the estimated fraction of the human genome available at the time of analysis, 66% (97,983/148,459) of SNPs currently match a BAC clone; this proportion will rise as human genomic sequence accumulates throughout this year. Mapped SNPs are placed freely in the public domain at (<http://snp.cshl.org/> and <http://www.ncbi.nlm.nih.gov/> SNP).

To verify the quality of candidate SNPs, we re-sequenced 183 loci in each of the individuals used in library construction. For 6 out of 183 (3%) loci, all individuals appeared 'heterozygous', indicating that the sequence variation was between several copies of a repetitive locus, rather than between different individual's copies of an orthologous locus. These will soon be easily eliminated by comparison with the full genome sequence. At the remaining 177 loci, 95% (216/227) of candidate SNPs were confirmed as polymorphic, demonstrating the accuracy of the techniques. We also validated 22 out of 24 (92%) candidate SNPs discovered solely by genomic alignment of RRS reads. Independent efforts have confirmed the high accuracy and heterozygosity of TSC SNPs. Using single-base extension to type over 1,000 TSC SNPs in the TSC discovery panel, essential identical validation results were obtained (M. Boyce-Jacino, personal communication). Furthermore, more than 80%

Table 1 Summary of SNP discovery

Total reads passing quality filters	653,348
Failed owing to high repeat content	123,993 (19%)
Passing reads	529,355 (81%)
Twins	528,937
Paired bases examined for SNPs	182,810,357
Heterozygous positions observed	142,651
Heterozygosity (π)	0.00078
Cliques	95,959
Cliques passing all quality filters	81,708
Unique SNPs identified	47,172
SNPs per passing read	1/11.2

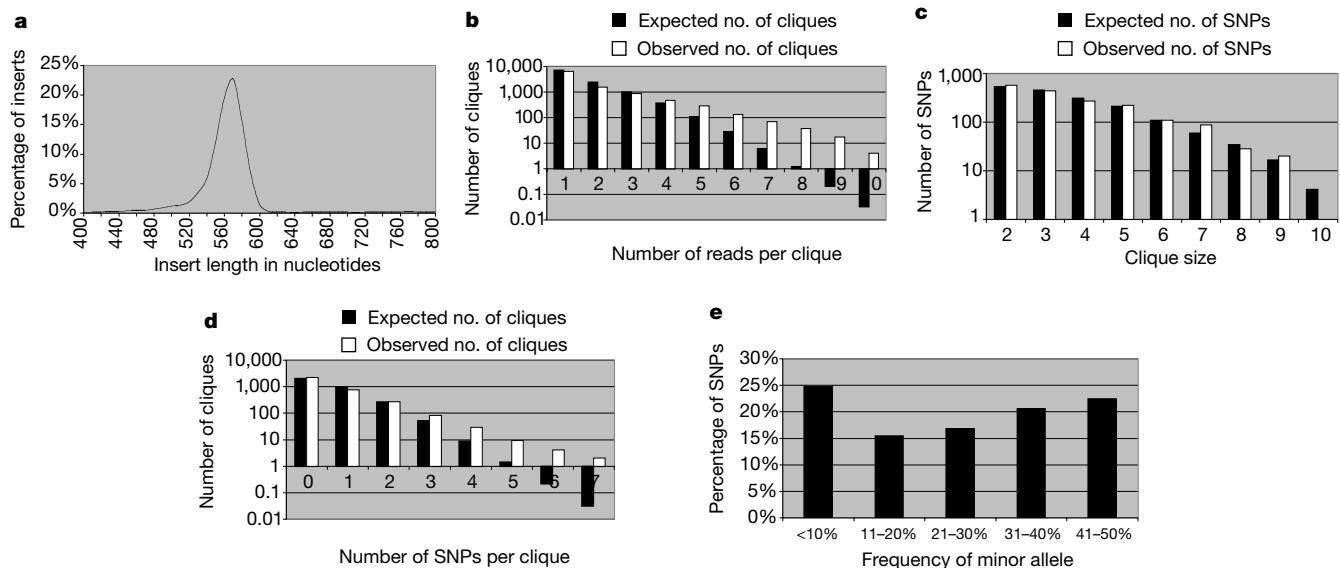


Figure 2 Pilot project data analysis. **a**, Insert size distribution. Lengths of inserts in a pilot RRS library cluster tightly around a mean of 560 bp, with 85% of reads between 535 and 585 bp. **b**, Clique size distribution. The number of cliques of each size was similar to that expected under composite Poisson sampling, with an excess of large cliques. These oversized cliques probably represent repetitive loci and were excluded from further analysis. **c**, Distribution of SNPs among cliques of different sizes. The number of SNPs found in cliques of different sizes is consistent with expectations under the neutral theory of evolution, assuming a constant rate of nucleotide diversity, neutral alleles and constant-sized population²². **d**, Distribution of loci containing different numbers of SNPs. The

number of loci with 0–7 SNPs closely matches the parameters described in **c**, with the exception of a few loci with an unexpectedly large number of SNPs. These may be attributable to inclusion of a small number of repeat sequences, or to varied levels of nucleotide diversity throughout the genome. **e**, Allele-frequency distribution of 214 validated SNPs among DNA samples used to construct the libraries (excluding two chromosomes in which the SNP was discovered). As expected, SNP discovery in a small number of chromosomes results in an allele frequency distribution that is relatively flat and thus yields a preponderance of common (>10%) alleles.

of more than 1,000 TSC SNPs were found to display high-frequency polymorphism (minor frequency > 5–10%) in independent population samples (P. Kwok, personal communication). Finally, the distribution of allele frequencies for validated SNPs in the diversity panel was roughly uniform, with a mean frequency of 24% (Fig. 2e).

We have demonstrated a new method for re-sampling loci without PCR, and for detecting SNPs with high accuracy in the resulting alignments. Although we focused on the human, RRS can be configured for any organism and scale, without the requirement for genome sequence data. Thus, RRS should be particularly useful for experimental models and agricultural species. In the human, the TSC's continuing discovery of SNPs will soon provide a map with average spacing of 3–5 kb, adequate for experimental characterization of linkage disequilibrium and haplotype-based disease association. These studies will require genotyping technology on a similar scale. In this regard, SNPs discovered by RRS offer a unique advantage: the potential for reduced representation genotyping (RRG) without locus-specific amplification. We are currently developing methods to isolate the reduced representation from each individual, generically amplify after linker ligation, and directly genotype the resulting fragments by established methods such as single-base extension¹⁹ or array hybridization¹⁴. When combined with a dense SNP map, such large-scale genotyping methods should markedly enhance attempts to unravel the genetic contribution to common diseases.

Note added in proof: Since submitting this manuscript, the number of SNPs discovered by the SNP Consortium has grown to over 350,000, of which 250,000 have been mapped to the draft sequence and are freely available at the TSC website (<http://snp.cshl.org>). □

Methods

Expected sampling of reduced representations

Given an expected average fragment size d , and a genome of size G , the number of unique

fragments (D) in the size range $\{x_1, x_2\}$ is estimated as:

$$D = (G/d)(e^{-x_1/d} - e^{-x_2/d})$$

With N reads performed from a library of D equally represented inserts, $N^2/2D$ pairwise matches are expected. However, laboratory separations result in bell-shaped size distributions (Fig. 2a), such that the chance any insert matches another depends on where it falls in the size distribution. The expected proportion of matches is then calculated as a composite Poisson distribution after dividing the data into appropriate size bins.

Library construction and sequencing

For the RRS pilot, an equimolar mixture of 10 DNAs was created (Coriell nos 10965 (Amerindian), 10470 (African Pygmy), 11322 (Chinese), 11589 (Japanese), 13820 (Russian), 05987A (Amish), 12615 (French), 11997 (Utah), 08779A (African-American), 10540 (Melanesian)) and digested to completion with *Bgl*II (New England Biolabs). Pooled, digested DNA (100 µg) was electrophoresed using 1.2% NuSieve agarose (BioWhittaker) and a narrow band (2 mm) was excised adjacent to a 564-bp size marker (*Lambda*/HindIII). We cloned fragments into M13mp19 RFI DNA (Pharmacia) and prepared clones by standard methods. We carried out dye-primer sequencing with the M13–21 forward primer and ABI 377 sequencers. For TSC production, we used the NIH diversity panel of 24 anonymous DNAs¹⁸. Each DNA (15 µg) was digested with *Bgl*II, *Hind*III, *Xba*I or *Eco*RI and pooled. Fragments were size separated on 1.6% SeaKemLE (BioWhittaker) agarose gels, excised in five adjacent size windows from 400 to 650 bp, and cloned into M13mp18. Sequences were obtained using standard methods and dye-primer chemistry and ABI 377 (96,960), dye-terminator chemistry and ABI 377 (22,165), or dye-terminator chemistry and ABI 3700 capillary sequencers (791,455).

Sequence alignment

We processed traces with PHRED, removed vector sequences and applied quality filters: we rejected reads with fewer than 100 NQS bases or inserts shorter than 200 bp. Passing sequences were compared by BLAST²⁰ to a library of known repeats (available on request) and rejected if more than 50% of bases matched known repeats. The remaining 'quality reads' were compared by BLAST, and pairs with > 80% identity over two-thirds of their length identified. Smith–Waterman alignment of overlapping sequence was followed by trimming of poor quality alignments at either end; if perfect alignment (10/10) did not begin within 50 bp of both ends, the pair was rejected. We identified sequence discrepancies meeting NQS and counted them. If more than 1% of NQS bases appeared polymorphic, the pair was rejected, as this level of diversity is uncommonly observed in comparison of orthologous loci from different individuals (Table 1; and refs 13–16, 21). Pairs passing these rules were clustered into cliques using transitivity. We compared clique sizes with the expected distribution under composite Poisson sampling for single-copy

loci in each length bin. Unexpectedly large cliques (with a less than 50% chance of being unique by comparison with expectation) were withheld.

SNP identification

A base met the NQS if its PHRED quality score was ≥ 20 , and the 5 bases to either side displayed PHRED scores ≥ 15 . Positions within alignments were considered high quality if both bases met the NQS and at least nine of the ten flanking base pairs were perfect matches. These values have been evaluated by detailed parametric testing, which will be presented elsewhere (V.J.P., B.V.E., D.A., E.S.L., unpublished data). To validate the rules, three finished BAC sequencing projects from the Whitehead Sequencing Center were selected (Genbank accession nos AC003950.1, AC007066 and AC004584, AC007159). External validation shows that the error rates of such finished sequences are less than 1 in 10,000 (C. Nusbaum, personal communication). Individual reads contributing to each BAC assembly were aligned to the consensus as described^{8,9}, and apparent discrepancies counted among positions with a PHRED quality (Q) score > 20 . Because our initial focus was in discovering SNPs, rather than insertion/deletions, we counted only substitutions. For this reason, the observed error rate is lower than that predicted by the PHRED score, which includes all classes of sequence errors (see refs 8, 9). SNP identification was fully automated using the rules above; no human revision was allowed. Similar results were observed for both dye-terminator and dye-primer chemistry, and with both slab-gel and capillary sequence detectors (data not shown).

SNP validation

Loci containing candidate SNPs were amplified by PCR from each of the DNA samples used to make the RRS libraries (10 DNAs for the pilot, 24 for the subsequent libraries and genomic validations), and sequenced according to standard methods. A repeat locus was declared if all individuals appeared heterozygous at one or more positions. Of unique loci, the candidate polymorphism was considered validated if two or three unambiguous, distinguishable genotypes were observed. For SNPs discovered by alignment to finished genomic sequence, we did not have access to DNA from the individual used to construct the BAC library, and instead used the same panel of 24 individuals. Given that some SNPs are rare (see Fig. 2e), we estimate that 5–10% of true SNPs would appear monomorphic in such a sample based on sampling variation alone.

Received 9 March; accepted 19 July 2000.

1. Lander, E. S. The new genomics: global views of biology. *Science* **274**, 536–539 (1996).
2. Collins, F. S., Guyer, M. S. & Chakravarti, A. Variations on a theme: cataloging human DNA sequence variation. *Science* **278**, 1580–1581 (1997).
3. Risch, N. & Merikangas, K. The future of genetic studies of complex human diseases. *Science* **273**, 1516–1517 (1996).
4. Hastbacka, J. *et al.* Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland. *Nature Genet.* **2**, 204–211 (1992); *erratum ibid.* **2**, 343 (1992).
5. Kruglyak, L. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nature Genet.* **22**, 139–144 (1999).
6. Collins, A., Lonjou, C. & Morton, N. E. Genetic epidemiology of single-nucleotide polymorphisms. *Proc. Natl Acad. Sci. USA* **96**, 15173–15177 (1999).
7. Weber, J. L. & Myers, E. W. Human whole-genome shotgun sequencing. *Genome Res.* **7**, 401–409 (1997).
8. Ewing, B. & Green, P. Base-calling of automated sequencer traces using PHRED. II. Error probabilities. *Genome Res.* **8**, 186–194 (1998).
9. Ewing, B., Hillier, L., Wendt, M. C. & Green, P. Base-calling of automated sequencer traces using PHRED. I. Accuracy assessment. *Genome Res.* **8**, 175–185 (1998).
10. Gu, Z., Hillier, L. & Kwok, P. Y. Single nucleotide polymorphism hunting in cyberspace. *Hum. Mutat.* **12**, 221–225 (1998).
11. Marth, G. T. *et al.* A general approach to single-nucleotide polymorphism discovery. *Nature Genet.* **23**, 452–456 (1999).
12. Buetow, K. H., Edmonson, M. N. & Cassidy, A. B. Reliable identification of large numbers of candidate SNPs from public EST data. *Nature Genet.* **21**, 323–325 (1999).
13. Cargill, M. *et al.* Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nature Genet.* **22**, 231–238 (1999).
14. Wang, D. G. *et al.* Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* **280**, 1077–1082 (1998).
15. Halushka, M. K. *et al.* Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. *Nature Genet.* **22**, 239–247 (1999).
16. Li, W. H. & Sadler, L. A. Low nucleotide diversity in man. *Genetics* **129**, 513–523 (1991).
17. Mullikin, J. C. *et al.* An SNP map of human chromosome 22. *Nature* **407**, 516–523 (2000).
18. Collins, F. S., Brooks, L. D. & Chakravarti, A. A DNA polymorphism discovery resource for research on human genetic variation. *Genome Res.* **8**, 1229–1231 (1998).
19. Landegren, U., Nilsson, M. & Kwok, P. Y. Reading bits of genetic information: methods for single-nucleotide polymorphism analysis. *Genome Res.* **8**, 769–776 (1998).
20. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
21. Cambien, F. *et al.* Sequence diversity in 36 candidate genes for cardiovascular disorders. *Am. J. Hum. Genet.* **65**, 183–191 (1999).
22. Li, W. H. *Molecular Evolution* (Sinauer Associates, Canada, 1997).

Acknowledgements

We are indebted to the staff of the Whitehead Institute/MIT Center for Genome Research Sequencing Center for high-throughput sequencing and to N. Stange-Thomann for contributions to library construction. We would like to thank B. Blumenstiel and R. Lane for library construction and SNP validation, and M. Molla, L. Friedland, J. Ireland and B. Gilman for informatics assistance. We appreciate helpful discussions with members of

The SNP Consortium, as well as colleagues at the Whitehead/MIT Genome Center. D.A. is a recipient of a Howard Hughes Medical Institute Postdoctoral Fellowship for Physicians. C.R.C. is supported by the Cancer Research Fund of the Damon Runyon / Walter Winchell Foundation. This work was conducted under grants from the Wellcome Trust and The SNP Consortium to E.S.L.

Correspondence and requests for materials should be addressed to E.S.L. (e-mail: lander@genome.wi.mit.edu).

An SNP map of human chromosome 22

J. C. Mullikin, S. E. Hunt, C. G. Cole, B. J. Mortimore, C. M. Rice, J. Burton, L. H. Matthews, R. Pavitt, R. W. Plumb, S. K. Sims, R. M. R. Ainscough, J. Attwood, J. M. Bailey, K. Barlow, R. M. M. Bruskewich, P. N. Butcher, N. P. Carter, Y. Chen, C. M. Clee, P. C. Coggill, J. Davies, R. M. Davies, E. Dawson, M. D. Francis, A. A. Joy, R. G. Lambie, C. F. Langford, J. Macarthy, V. Mall, A. Moreland, E. K. Overton-Larty, M. T. Ross, L. C. Smith, C. A. Steward, J. E. Sulston, E. J. Tinsley, K. J. Turney, D. L. Willey, G. D. Wilson, A. A. McMurray, I. Dunham, J. Rogers & D. R. Bentley

The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

The human genome sequence will provide a reference for measuring DNA sequence variation in human populations. Sequence variants are responsible for the genetic component of individuality, including complex characteristics such as disease susceptibility and drug response. Most sequence variants are single nucleotide polymorphisms (SNPs), where two alternate bases occur at one position^{1–3}. Comparison of any two genomes reveals around 1 SNP per kilobase^{1,3}. A sufficiently dense map of SNPs would allow the detection of sequence variants responsible for particular characteristics on the basis that they are associated with a specific SNP allele^{4–6}. Here we have evaluated large-scale sequencing approaches to obtaining SNPs, and have constructed a map of 2,730 SNPs on human chromosome 22. Most of the SNPs are within 25 kilobases of a transcribed exon, and are valuable for association studies. We have scaled up the process, detecting over 65,000 SNPs in the genome as part of The SNP Consortium programme, which is on target to build a map of 1 SNP every 5 kilobases that is integrated with the human genome sequence and that is freely available in the public domain.

Single nucleotide polymorphisms (SNPs) are stable, bi-allelic sequence variants that are distributed throughout the genome, which can be assayed using high-throughput automated methods. Sequence variants have been detected previously by analysis of sequence differences in clusters of expressed sequence tags^{7,8}; or by re-sequencing DNA fragments after amplification from different individuals^{9–11}, sometimes following prescreening^{12–14}. These approaches are effective for exploring sequence variation of individual genes in depth. An alternative approach, which takes advantage of the reagents from the human genome project, is to detect SNPs in regions of overlap between bacterial clones containing sequences of independent genomes (ref. 2; and E.D. *et al.*, unpublished data). This analysis provides a valuable resource of SNPs in short sections throughout the genome, but each section is interspersed by large regions (typically 0.1–0.5 megabases (Mb)) where the sequence of only one genome is available, and where no polymorphisms can be detected.

We evaluated two large-scale sequencing strategies to identify SNPs to cover the entire human genome at a target density of 1 SNP per 5 kilobases (kb). In the reduced representation shotgun (RRS)

strategy¹⁵, specific subsets of restriction fragments, made from an equimolar mixture of DNA isolated from unrelated individuals, are repeatedly sampled by sequencing. Reads derived from the same fragment in different genotypes are aligned into clusters, or cliques, and high-confidence sequence differences between any two reads (that is, candidate SNPs) are recorded. Experimental verification of a subset of these candidate SNPs (by re-sequencing to identify the SNP in individual samples of the original DNA panel) tests the criteria used in the computational detection of candidate SNPs. In the genomic-alignment strategy, single reads obtained by shotgun sequencing of a library of DNA fragments are aligned directly to available genomic sequence to detect the candidate SNPs. The two strategies are complementary: the RRS strategy allows detection of SNPs throughout the genome without genomic sequence, but the SNPs are not automatically mapped; whereas the genomic-alignment strategy requires genomic sequence, and provides a map location for each SNP. Both strategies are applicable to any genome.

The availability of the complete sequence of chromosome 22 (ref. 16) enabled us to compare both approaches directly. The efficiency of SNP detection for each approach was measured as the number of SNPs detected relative to the amount of new raw sequence data generated for the analysis (see Table 1). For RRS, chromosome 22 was flow-sorted from seven unrelated individuals (see Methods), and clones from a library of 1.2–2.1-kb fragments generated by *HindIII* digestion were sequenced. After removing highly repetitive or poor quality sequences, 4,584 reads were aligned into 1,842 clusters (1,013 with more than one read, plus 829 singletons), corresponding to a mean cluster depth (number of reads in clusters / number of clusters) of 2.49. Using a specific set of criteria (see Methods), we detected 455 candidate SNPs in the 1,013 clusters of two or more reads. Experimental verification of a subset of these candidates by re-sequencing confirmed that 95% (74/78) were true SNPs. The remaining four candidate SNPs were homozygous in all DNA samples tested, and are presumed to be sequencing errors. The efficiency of SNP detection was 1 SNP per 10.1 reads (1 SNP per 4.79 kb of raw data; see Table 1). In a separate analysis we detected an additional 13% variants comprising insertion/deletion polymorphisms. Two-thirds of these were variations in poly(A) tracts, and the remainder (4% of all variants) are potentially a valuable additional source of polymorphisms for genetic studies.

In contrast to the RRS strategy, in which the library must be sequenced to a sufficient depth to obtain clusters of multiple reads for SNP detection, genomic-alignment analysis minimally requires alignment of just one read against finished genomic sequence. The genomic-alignment strategy should therefore result in a higher efficiency of SNP detection. Like RRS, the genomic-alignment strategy should also yield high-confidence SNPs, as all human genomic sequence is finished to an accuracy of more than 99.99%, (and therefore has a quality value *Q* of 40; ref. 17). To test this hypothesis, we aligned the RRS reads obtained from the previous experiment (including all singletons) to the finished

chromosome 22 sequence (33.4 Mb). We identified 914 candidate SNPs, and included all the SNPs found by the RRS analysis. The success rate of verification was 94% (115/122). The strategy therefore results in a twofold improvement in the efficiency of SNP detection (1 SNP per 5.0 reads compared with 1 SNP per 10.1 reads; see Table 1).

The genomic-alignment approach should be most efficient if each new sequence read aligns to a separate section of the genome. We therefore constructed a library of randomly sheared fragments from flow-sorted chromosome 22 DNA, and aligned 5,567 high-quality reads to the finished genomic sequence. We detected 1,845 candidate SNPs, and the verification success rate was 97% (33/34). The SNP detection rate in this analysis is 1 SNP per 1,391 bases of raw sequence data of *Q* ≥ 23 (or 1 SNP per 3.0 reads). Genomic-alignment analysis of the random shotgun sequences (compared with genomic-alignment analysis of the RRS sequences) thus provided a further 1.6-fold improvement in the efficiency of SNP detection.

A total of 2,730 different SNPs (that is, 1 per 12 kb of chromosome 22 sequence) were identified during these studies. The position of each SNP was determined by aligning its flanking sequence to the sequence of chromosome 22. The plots in Fig. 1 illustrate the distribution of SNPs detected by each method. Chromosome 22q contains at least 545 transcribed genes (3,632 exons) on the basis of the reported annotation. 1,043 of the SNPs (38%; magenta bars in Fig. 1) detected in this study lie either inside or within 5 kb of a transcribed exon in the current annotated set. In all, 1,771 (65 %) of the SNPs are within 25 kb of an exon, and may be informative in association studies, depending on the extent of linkage disequilibrium in genomic regions^{5,18,19}. From the current set of annotated transcribed exons, 37% (1,333) of them have at least one SNP within 5 kb, and 84% (3,039) have at least one SNP from the present set within 25 kb. This study therefore already provides SNPs that are sufficiently close to most of the transcribed regions to be in linkage disequilibrium with possible functional variants, and this coverage will improve when a more dense SNP map is produced for the whole genome.

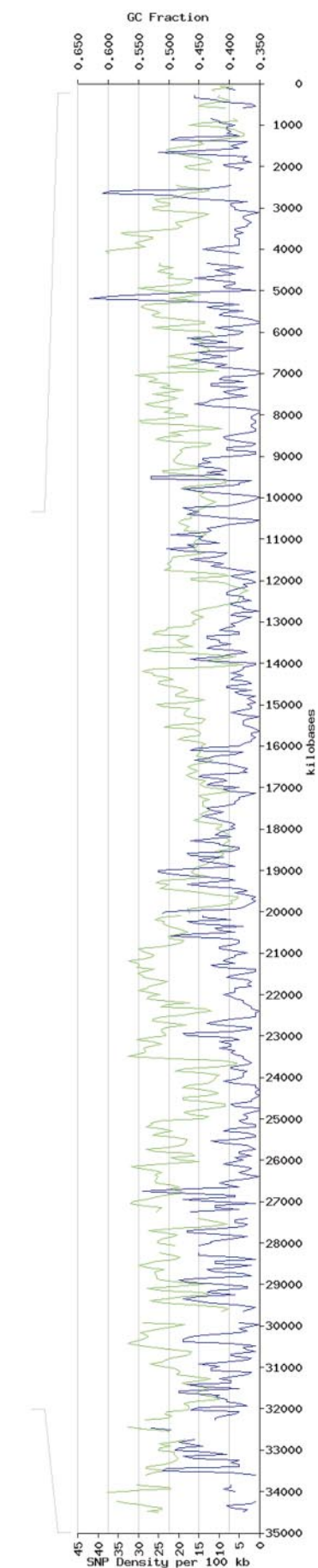
On the basis of the results of this study, and studies at the Whitehead Institute and the Genome Sequencing Center, St Louis, The SNP Consortium (TSC: a consortium of academic and pharmaceutical companies, see <http://snp.cshl.org>) initiated a programme with the goal of generating a freely available public resource of 300,000 SNPs, of which at least 150,000 would be mapped by April 2001. As part of this work, we have extended our SNP identification programme by application of the RRS strategy to a whole-genome library of *PvuII* fragments of 0.925–1.250 kb, constructed using DNA isolated from 24 unrelated individuals (the same panel used by all TSC participants; see Methods). Up to April 2000, sequencing to a cluster depth of 2.51 has yielded 52,354 candidate SNPs. This corresponds to a detection rate of 1 SNP per 9.0 reads. The verification success rate for these SNPs was 97% (122/128). Candidate SNPs in three additional loci were heterozygous in all DNA samples tested, and these loci are presumed to be low-copy repeats. The success rate of SNP verification concurs with the results of the pilot study, and confirms the feasibility of the RRS approach to detect SNPs in the whole genome. Furthermore, alignment of the *PvuII* RRS data with the 500 Mb finished genomic sequence that is available resulted in identification of a total 19,192 SNPs (including 9,727 not detected by RRS). On the basis that roughly 2,700 Mb of the genome sequence becomes available, the number of SNPs found by genomic-alignment analysis of this set of RRS reads would be twofold higher than by RRS analysis alone, in agreement with the results of the chromosome 22 study.

Production of the working draft sequence (comprising mapped bacterial clones each sequenced to a depth of ≥ threefold in bases of *Q* ≥ 20; ref. 20) of the human genome is well advanced (<http://www.nhgri.nih.gov/HGP>). Genomic-alignment analysis of RRS or

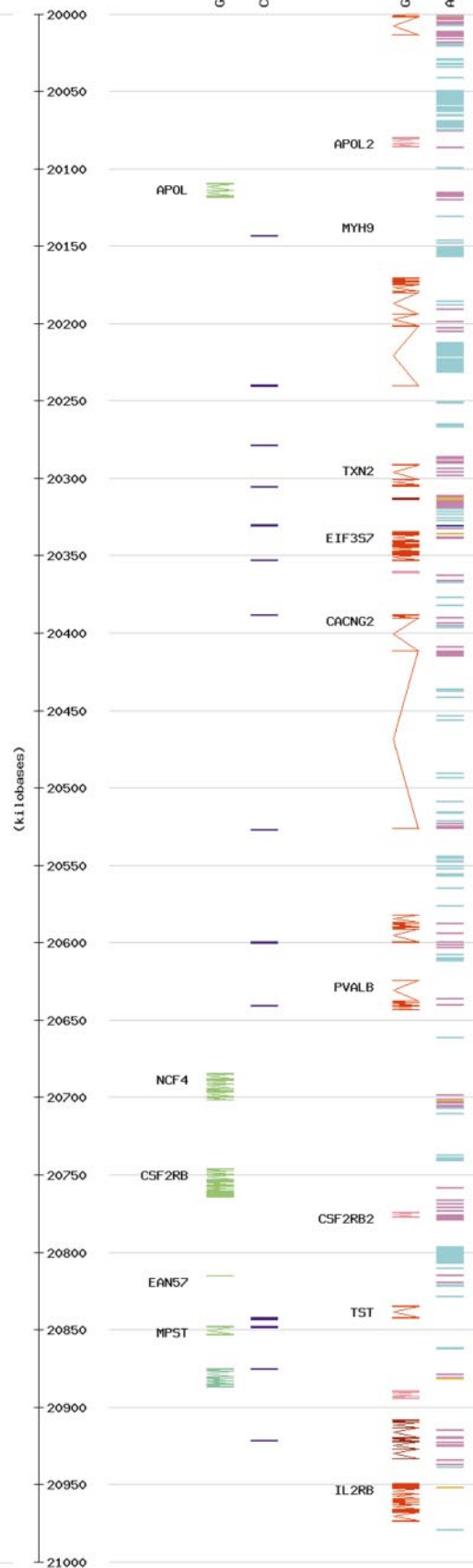
Table 1 Comparison of strategies for SNP identification

	Chromosome 22			Whole genome
	<i>HindIII</i> RRS	RRS-GA	Random GA	<i>PvuII</i> RRS
SNPs detected	455	914	1,845	52,354
Reads	4,584	4,584	5,567	473,249
Clusters	1,842	1,842	5,401	188,545
Cluster depth	2.49	3.49	2.03	2.51
SNPs per kb raw data	1/4.79	1/2.38	1/1.39	1/3.98
SNPs per read	1/10.1	1/5.0	1/3.0	1/9.0

Summary of SNP detection by different strategies. Clusters, the number of unique genomic loci represented in the clusters obtained (including those with single reads); cluster depth, (number of reads in clusters)/(number of clusters); genomic sequence is counted as a single read per cluster where applicable (columns 2 and 3); SNPs per read, SNPs detected per read analysed; SNPs per kb raw data, an exact measure of relative efficiency, in SNPs per kilobase (at *Q*23) of raw data at *Q*23 used in the analysis. The total amount of *Q*23 raw sequence data used in each analysis was 2.18 Mb, 2.18 Mb, 2.57 Mb and 208.2 Mb, respectively. GA, genomic alignment.



Density Plot
 ■ GC Content
 ■ SNP Density



Genes
 Transcript(+) ■
 Transcript(-) ■
 Known gene ■
 Related gene ■
 Predicted gene ■

All SNPs Colour Coding
 ■ Inside Exon ■ Within 5 kb ■ Inside CpGI
 ■ SNPs, not in above categories



random shotgun sequences using the draft would provide a high yield of candidate SNPs. The sequence reads are assembled using the program PHRAP (P. Green, personal communication), which provides a combined quality score for each base of the consensus sequence. This PHRAP quality score can be used for identification of candidate SNPs by genomic-alignment analysis using the working draft, in the same way as using finished sequence, by requiring a minimum PHRAP quality score (Q) of 40. After aligning the sequence data from the *PvuII* RRS library against a sample of unfinished sequence with variable quality scores, a set of candidate SNPs was selected using a PHRAP Q value of ≥ 40 . The verification success rate was 99% (244/246). This confirmed the validity of using genomic-alignment analysis with the working draft sequence of the genome to identify (and map) candidate SNPs.

The initial goal of the TSC was to identify 300,000 SNPs (1 per 10 kb) by April 2001. Our work shows that the goal can be achieved using the RRS strategy and the resources currently available in the consortium. To our benefit, the acceleration in the human genome sequencing programme over the past 12 months will provide the opportunity to perform genomic-alignment analysis on all the sequence data, thus providing a substantial improvement in the yield of SNPs (at least twofold on current projections, or 600,000 SNPs) from the TSC programme. Furthermore, the majority of SNPs will be mapped directly by alignment to the genomic sequence, thus providing an SNP map (1 SNP per 5 kb on average) of the human genome. This freely available, public resource will underpin extensive genetic studies to establish the extent of linkage disequilibrium in the human genome, and to investigate the association of genes with complex disease and other phenotypic traits. □

Methods

Sequencing and alignment

Chromosome 22 was flow-sorted from individual lymphoblastoid cell lines derived from seven individuals of north European origin, selected from the Porton Down collection of unrelated individuals collected as controls for disease association studies (Panel HRC, nos 159, 146, 184, 163, 160, 226, 193, 575 and 148). Chromosome preparations were pooled and digested with *HindIII*, and a gel-purified size fraction representing 1.2–2.1 kb was subcloned into the plasmid vector pUC18. We picked clones from a library of more than 100,000 transformants and sequenced them using forward and reverse primers as described²¹. Data were collected on ABI377 or 3700 machines (PE Biosystems), and analysed using the base-calling algorithm PHRED^{22,23}. All sequence reads that contained more than 100 bases with a PHRED quality score (Q) ≥ 30 were selected for masking of repeats using RepeatMasker (version 21/03/99: A. Smit and P. Green, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>). Reads with more than 80 bases of unique sequence were then assembled using a method based on Cross_Match followed by multiple sequence alignment of the reads matching unique genomic loci on chromosome 22.

The *PvuII* RRS libraries were constructed using genomic DNA from 24 individuals containing one or more representatives of a range of ethnic groups (the DNA Polymorphism Discovery Resource M24PDR, Coriell Cell Repositories²⁴), by agarose gel fractionation of a complete digest of a mixture of the DNAs, followed by subcloning of specific size fractions in pUC18. For assembly of forward and reverse sequence reads, PHRAP (P. Green, University of Washington) was used to assemble the RRS *PvuII* library data (see Table 1).

Candidate SNPs

Criteria for selection of candidate SNPs were as follows (see also ref. 15): (1) the quality value (Q) of the SNP base (a value recorded for each base in a sequence read by the base-calling program PHRED^{22,23}) was ≥ 23 ; and the Q value for the 5 bases on either side of the

SNP was ≥ 15 . (2) At least nine of the flanking ten bases matched between reads. All ten had to match for the Cross_Match (P. Green, University of Washington) based method. (3) The cluster depth was no greater than eight reads, on the basis that deeper clusters might comprise a low-copy repeat. (4) The number of candidate SNPs in a cluster was ≤ 4 , on the basis that clusters with more divergent sequences might be composed of low-copy repeats (that is, recently diverged paralogous sequences, accumulating sequence differences between them). The minimum Q for the SNP base was chosen as follows: visual examination of the trace data for a randomly selected set of candidate SNPs called at Q ≥ 20 revealed 15 out of 111 (13%) obvious sequence artefacts, 9 of which had Q values in the range from 20 to 22. When the SNP identification was rerun at Q ≥ 23 , a new randomly selected set of candidate SNPs was examined, from which only 5 out of 117 (4.3%) failures were observed. This failure rate was confirmed by the experimental verification data.

Classification of single-base substitutions

We classified single-base substitutions on the basis of transitions or transversions as follows. C to T (or G to A) transitions: 70.1% of all SNPs were possible CpG to TpG mutations (the most frequently observed single-base substitution, which is presumed to arise following deamination of 5-methylcytosine in Me-CpG; ref. 25); 29.1% of all SNPs were transversions of C/A, C/G and T/A (15.7%, 8.6% and 5.6% of all SNPs, respectively).

SNP verification

Candidate SNPs from each dataset were selected at random for experimental verification. Primer pairs were designed using PRIMER (http://www.genome.wi.mit.edu/genome_software/genome_software_index.html) and loci were amplified from the DNA samples of the individual cell lines used for the initial library construction. Polymerase chain reaction products were purified by treatment with shrimp alkaline phosphatase and exonuclease I, sequenced from both ends, and the data for each SNP assembled and examined in a GAP4 database.

Data access

All data including sequence reads, candidate SNPs and verification information were submitted to the Data Coordination Centre (DCC) of The SNP Consortium (<http://snp.cshl.org>). Each SNP is assigned a unique identifier (for example, TSC0137673) and released in the public domain when given a map position either by alignment to mapped genomic sequence, or by whole-genome radiation hybrid mapping. SNP information is available at the above web site (from the home page, click on 'data', then 'object search', and enter 'TSC0137673', and search); and also in dbSNP (<http://www.ncbi.nlm.nih.gov/SNP>).

Received 10 March; accepted 19 July 2000.

- Wang, D. G. *et al.* Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* **280**, 1077–1082 (1998).
- Taillon-Miller, P., Gu, Z., Li, Q., Hillier, L. & Kwok, P. Y. Overlapping genomic sequences: a treasure trove of single-nucleotide polymorphisms. *Genome Res.* **8**, 748–754 (1998).
- Kwok, P. Y., Deng, Q., Zakeri, H., Taylor, S. L. & Nickerson, D. A. Increasing the information content of STS-based genome maps: identifying polymorphisms in mapped STSs. *Genomics* **31**, 123–126 (1996).
- Risch, N. & Merikangas, K. The future of genetic studies of complex human diseases. *Science* **273**, 1516–1517 (1996).
- Collins, F. S., Guyer, M. S. & Chakravarti, A. Variations on a theme: cataloging human DNA sequence variation. *Science* **278**, 1580–1581 (1997).
- Chakravarti, A. Population genetics—making sense out of sequence. *Nature Genet.* **21**, 56–60 (1999).
- Buetow, K. H., Edmonson, M. N. & Cassidy, A. B. Reliable identification of large numbers of candidate SNPs from public EST data. *Nature Genet.* **21**, 323–325 (1999).
- Picoult-Newberg, L. *et al.* Mining SNPs from EST databases. *Genome Res.* **9**, 167–174 (1999).
- Nickerson, D. A. *et al.* DNA sequence diversity in a 9.7-kb region of the human lipoprotein lipase gene. *Nature Genet.* **19**, 233–240 (1998).
- Cargill, M. *et al.* Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nature Genet.* **22**, 231–238 (1999).
- Halushka, M. K. *et al.* Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. *Nature Genet.* **22**, 239–247 (1999).
- Orita, M., Iwahana, H., Kanazawa, H., Hayashi, K. & Sekiya, T. Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proc. Natl Acad. Sci. USA* **86**, 2766–2770 (1989).
- Ganguly, A., Rock, M. J. & Prockop, D. J. Conformation-sensitive gel electrophoresis for rapid detection of single-base differences in double-stranded PCR products and DNA fragments: evidence for solvent-induced bends in DNA heteroduplexes. *Proc. Natl Acad. Sci. USA* **90**, 10325–10329 (1993).
- O'Donovan, M. C. *et al.* Blind analysis of denaturing high-performance liquid chromatography as a tool for mutation detection. *Genomics* **52**, 44–49 (1998).
- Altshuler, D. *et al.* An SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature* **407**, 513–516 (2000).
- Dunham, I. *et al.* The DNA sequence of human chromosome 22. *Nature* **402**, 489–495 (1999).
- Felsenfeld, A., Peterson, J., Schloss, J. & Guyer, M. Assessing the quality of the DNA sequence from the Human Genome Project. *Genome Res.* **9**, 1–4 (1999).
- Collins, A., Lonjou, C. & Morton, N. E. Genetic epidemiology of single-nucleotide polymorphisms. *Proc. Natl Acad. Sci. USA* **96**, 15173–15177 (1999).
- Kruglyak, L. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nature Genet.* **22**, 139–144 (1999).
- Bouck, J., Miller, W., Gorrell, J. H., Muzny, D. & Gibbs, R. A. Analysis of the quality and utility of random shotgun sequencing at low redundancies. *Genome Res.* **8**, 1074–1084 (1998).
- The Sanger Centre & The Washington University Genome Sequencing Centre. Toward a complete human genome sequence. *Genome Res.* **8**, 1097–1108 (1998).

Figure 1 Single nucleotide polymorphism map of human chromosome 22. The four columns of coloured bars represent the SNP map determined by each analysis. See Table 1 for SNP totals corresponding to the three columns 'RRS', 'RRS-GA' and 'Random GA'; the column 'All SNPs' contains the non-redundant set of 2,730 SNPs. GA, genomic alignment. Gold and magenta bars denote SNPs that are located inside or within 5 kb of an exon, respectively. One region of the chromosome (20,000–21,000 kb; coordinates as in ref. 16; see also <http://www.sanger.ac.uk/HGP/Chr22>) is enlarged to show the positions of individual SNPs relative to annotated exons and putative CpG islands.

22. Ewing, B., Hillier, L., Wendl, M. C. & Green, P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8**, 175–185 (1998).
23. Ewing, B. & Green, P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8**, 186–194 (1998).
24. Collins, F. S., Brooks, L. D. & Chakravarti, A. A DNA polymorphism discovery resource for research on human genetic variation. *Genome Res.* **8**, 1229–1231 (1998).
25. Holliday, R. & Grigg, G. W. DNA methylation and mutation. *Mutat. Res.* **285**, 61–67 (1993).

Acknowledgements

We are grateful for the contribution of all the members of the sequencing and supporting teams, and members of the informatics group of the Sanger Centre (see <http://www.sanger.ac.uk>). We thank the Wellcome Trust and The SNP Consortium for financial support, and members of the participating centres for helpful discussions.

Correspondence and requests for materials should be addressed to D.R.B. (e-mail: drb@sanger.ac.uk).

A vertebrate globin expressed in the brain

Thorsten Burmester^{†‡}, Bettina Weich[‡], Sigrid Reinhardt[§]
& Thomas Hankeln^{†‡}

^{*} Institutes of Zoology, [‡] Molecular Genetics, Biosafety Research and Consulting, and [§] Physiological Chemistry, Johannes Gutenberg University Mainz, D-55099 Mainz, Germany

[†] These authors contributed equally to this work

Haemoglobins and myoglobins constitute related protein families that function in oxygen transport and storage in humans and other vertebrates^{1,2}. Here we report the identification of a third globin type in man and mouse. This protein is predominantly expressed in the brain, and therefore we have called it neuroglobin. Mouse neuroglobin is a monomer with a high oxygen affinity (half saturation pressure, $P_{50} \approx 2$ torr). Analogous to myoglobin, neuroglobin may increase the availability of oxygen to brain tissue. The human neuroglobin gene (*NGB*), located on chromosome 14q24, has a unique exon–intron structure. Neuroglobin

represents a distinct protein family that diverged early in metazoan evolution, probably before the Protostomia/Deuterostomia split.

Globins are porphyrin-containing proteins that bind oxygen reversibly and are therefore important in the respiratory system of living species¹. They have been found in many taxa, including bacteria, plants, fungi and animals³. Two types of globins have been described in vertebrates. The heterotetrameric haemoglobins transport oxygen in the blood, whereas the monomeric myoglobin of muscle cells facilitates the diffusion of oxygen to the mitochondria⁴. Although globins are among the best-investigated vertebrate proteins and several functional variants of the haemoglobin subunits are known¹, no other distinct types of globins have been identified so far in this taxon.

In the databases of anonymous mouse and human complementary DNAs (expressed sequence tags; ESTs⁵), we found partial globin-like sequences that do not correspond to any known haemoglobin or myoglobin. We cloned and sequenced the coding regions of the human and mouse cDNAs and the genomic region of the human gene. The mouse and human gene each code for proteins of 151 amino acids (relative molecular mass 17,000; M_r 17K) that are 94% identical; this is higher than the conservation between the orthologous haemoglobins or myoglobins of these species (77–85% identity) and within the uppermost range of more than 1,100 proteins compared between man and mouse⁶. Although the proteins clearly belong to the globin superfamily, they share little amino-acid sequence similarity with vertebrate myoglobins (< 21% identity) and haemoglobins (< 25% identity), suggesting a distinct evolution and function (Fig. 1).

We analysed the expression pattern of the human gene by northern hybridization to a filter containing RNA from different tissues and developmental stages (Table 1; see Supplementary Information). We note a predominant expression in the brain with the strongest signals observed in the frontal lobe, the subthalamic nucleus and the thalamus. Globin expression was also detected by messenger RNA *in situ* hybridization in neuronal cells of mouse brain regions (Fig. 2). We therefore propose to designate this protein neuroglobin (NGB). Polymerase chain reaction with reverse transcription (RT–PCR) experiments using murine RNA confirmed that other tissues contain only minor amounts of *Ngb*

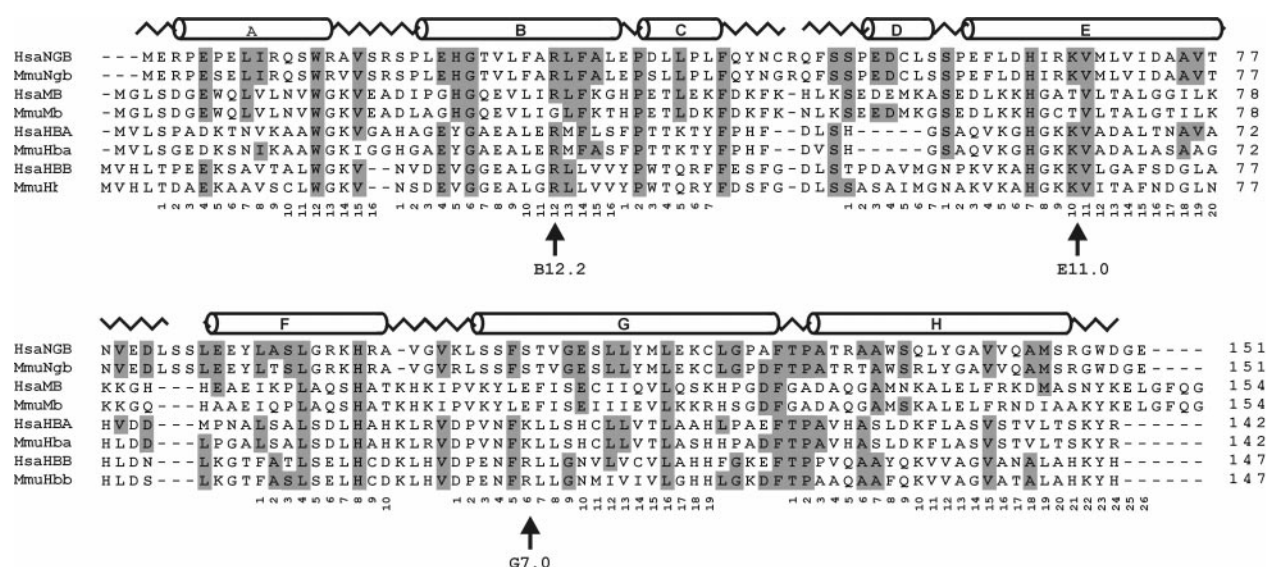


Figure 1 Comparison of human and mouse neuroglobin (HsaNGB and MmuNgb) with myoglobins (HsaMB, accession number M14603; MmuMb, P04247) and haemoglobins α and β (HsaHBA, J00153; HsaHBB, M36640; MmuHba, A45964; MmuHbb, P02088). The globin consensus numbering is given below the sequences, the secondary structure of the

human haemoglobin β is superimposed in the upper row. α -Helices are designated A to H, amino acids conserved between the neuroglobins and the myoglobins or haemoglobins are shaded. Intron positions in the human NGB sequence (at B12.2, E11.0 and G7.0) are indicated by arrows.

mRNA (data not shown), either reflecting its expression in inner-vating neurons, or the presence of low concentrations of Ngb in the cells of these tissues.

We partially purified neuroglobin from mouse brain (see Supplementary Information). The concentration of the Ngb protein is low and estimated to be less than 0.01% of the total protein content. Size-exclusion chromatography shows that both native and recombinantly expressed mouse Ngb are monomers of about 16K (data not shown). Both proteins bind oxygen reversibly. The spectra of recombinant and native mouse Ngb are identical, with three maxima at 424, 528 and 558 nm in the deoxygenated form (red protein) and a single maximum at 413 nm as oxy-neuroglobin (yellow protein) (see Supplementary Information). Although the maxima in the ultraviolet region are similar to those of most other globins⁷, the spectrum in the visible range is unusual and indicates the presence of a low-spin haem group, as observed, for example, for the nerve globin of the mollusc *Spisula solidissima*⁸. Recombinant Ngb has an oxygen affinity (P_{50}) of 1.9 to 2.3 torr, which is higher than that of mammalian haemoglobin (~26 torr), but lower than that of myoglobin (~1 torr)^{1,2}.

Radiation hybrid mapping places the human *NGB* gene on chromosome 14q24 between markers D14S76 and WI-4643. The genomic sequence was obtained from a human (P1-derived artificial chromosome; PAC) clone. The promoter region contains several putative Sp1-binding sites and at least three transcription starting points (identified by 5' rapid amplification of cDNA ends (RACE); data not shown), but lacks a TATA box motif. These features indicate a housekeeping gene promoter. The *NGB* gene has three introns at positions B12-2 (that is, between codon positions 2 and 3 of the twelfth amino acid of helix B), E11-0 and G7-0 (see Fig. 1). The introns B12-2 and G7-0 in *NGB* are conserved in haemoglobins and myoglobins of vertebrates and many other taxa^{3,9}, but the central intron at E11-0 is unprecedented. On the basis of protein structural considerations, however, the presence of a

central intron in ancient globin genes was previously proposed to be exactly at this position¹⁰. Therefore, the vertebrate globin gene ancestor might have displayed a 3 intron/4 exon structure. Alternatively, the central E11-0 intron of neuroglobins may represent a case of independent intron acquisition^{11,12}.

Globins constitute a valuable model for investigating evolution at the molecular level¹³. Although the resolution of the phylogenetic tree is poor, evolutionary analysis shows that the neuroglobins do not group with vertebrate haemoglobins and myoglobins (Fig. 3). Haemoglobins and myoglobins of vertebrates are thought to have diverged about 550 million years ago (Myr)¹³, but the lineage leading to the vertebrate neuroglobins must be older. A distinct position of neuroglobins is also suggested by its unique exon-intron structure. There is weak but consistent support for an association of the neuroglobins with the intracellular globins of the annelids *Aphrodite aculeata*¹⁴ and *Glycera dibranchiata*¹⁵⁻¹⁷, which are phylogenetically distinct from the extracellular globins of this phylum¹⁴. The *Aphrodite* globin displays the highest similarity with the neuroglobins (30% amino-acid identity). Taking into account the nerve-based expression of both the *Aphrodite* globin and the vertebrate neuroglobins, both may be functionally and

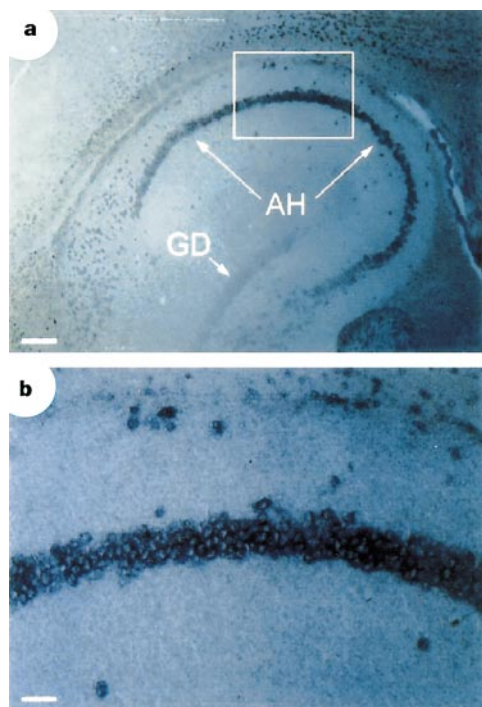


Figure 2 Neuroglobin expression in adult mouse brain (sagittal section) was studied by *in situ* hybridization using a digoxigenin-labelled oligonucleotide probe. **a**, Overview of the hippocampus formation with dentate gyrus (GD) and the C-shaped structure of the Ammon's horn (AH). **b**, Enlargement of framed part in **a** depicts an area of the Ammon's horn, where *Ngb* mRNA could be detected in the pyramidal cell layer.

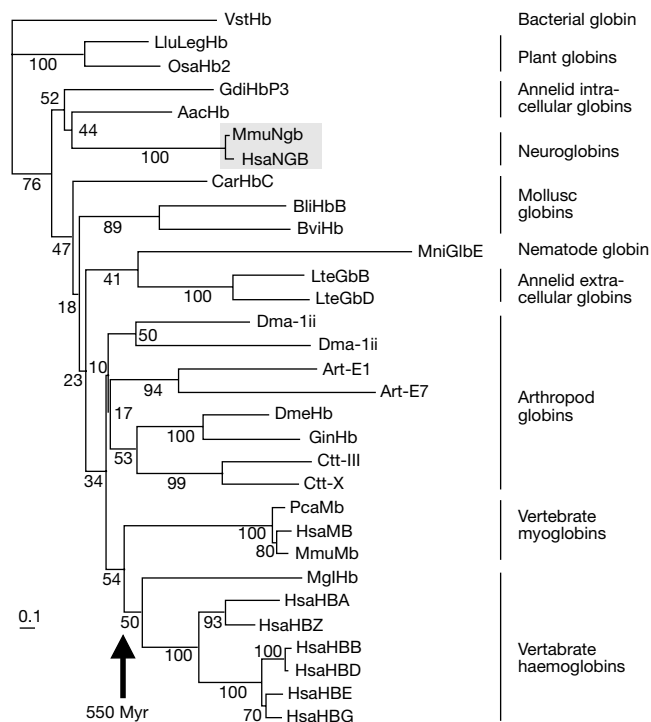


Figure 3 Neighbour-joining tree of selected globins. Bootstrap values are given at the branches; scale bar represents a distance of 0.1 accepted point mutations per site (PAM). The neuroglobins are shaded. We included other proteins in addition to those in Fig. 1. Bacteria: *Vitreoscilla stercoraria* haemoglobin (VstHb; P04252). Plants: *Lupinus luteus* leghaemoglobin II (LluLegHb; S41328); and *Oryza sativa* haemoglobin 2 (OsaHb2; U76028). Invertebrates: *Barbatia lima* haemoglobin β (BliHbB; D63934); *Barbatia virescens* haemoglobin (BviHb; D58416); *Aphrodite aculeata* nerve myoglobin (AachHb; U46754); *Glycera dibranchiata* P3 (GdiHbP3; P23216); *Lumbricus terrestris* globin B (LteGbB; P02218) and globin E (LteGbD; P08924); *Mermis nigrescens* eye globin (MniGlbE; AF138291); first (Dma1.1) and second domain (Dma1.2) of the haemoglobin (U67067) of *Daphnia magna*; *Artemia salina* globins E-1 (AsaE-1; P19364) and E-7 (AsaE7; P19363); *Gasterophilus intestinalis* haemoglobin (GinHb; AF063938); *Drosophila melanogaster* globin (DmeGlob1; AJ132818); *Chironomus thummi* haemoglobins Ctt-III (P02229) and Ctt-X (P02228). Vertebrates: *Myxine glutinosa* haemoglobin (MglHb; AF156936); *Physeter catodon* myoglobin (PcaMb; P02185); *Homo sapiens* haemoglobins γ (HsaHBG; M15386), δ (HsaHBD; P02042), ϵ (HsaHBE; I36933) and ζ (HsaHBZ; M24173).

Table 1 *NGB* expression in human tissues

Tissue	Expression level (%)
Brain	
Subthalamic nucleus	100
Frontal lobe	73
Thalamus	73
Occipital pole	70
Medulla oblongata	61
Temporal lobe	50
Cerebral cortex	44
Substantia nigra	43
Putamen	21
Amygdala	13
Caudate nucleus	11
Hippocampus	11
Cerebellum	10
Whole brain	23
Fetal brain	34
Other tissues	
Pituitary gland	62
Appendix	25
Adrenal gland	21
Lung	19
Colon	10

We quantified *NGB* expression in human tissues using an RNA Master Blot containing normalized amounts of poly(A)⁺ RNA. Less than 10% expression relative to the subthalamic nucleus was observed in spinal cord, pancreas, small intestine, stomach, testes, lymph node, ovary and thymus. *NGB* is not expressed in heart, aorta, skeletal muscle, bladder, uterus, prostate, thyroid gland, salivary gland, mammary gland, liver, spleen, peripheral leukocyte, lymph node, bone marrow, trachea or placenta, nor in the fetal heart, kidney, liver, spleen, thymus or lung.

phylogenetically homologous. Our analyses suggest that these nerve-based globins diverged from the other globins before Protostomia and Deuterostomia separated.

The presence of a third globin type in vertebrates is surprising because no specialized proteins were considered necessary to increase the intracellular oxygen pressure in tissues other than the muscle. However, nerve-based globins have been observed in invertebrate taxa such as annelids, molluscs and nematodes^{18–20}. In the mollusc *Tellina alternata*, the neural excitability depends on the oxygen stored by a nerve globin²¹. The oxygen-binding properties of neuroglobin are comparable to those of a typical vertebrate myoglobin, suggesting a similar function of neuroglobin in the brain. The supply of the brain with adequate amounts of oxygen is essential: although it makes up only 2% of the body weight, the brain consumes about 20% of the available oxygen during rest²². Despite of its low cellular concentration, the high oxygen affinity of neuroglobin may help to transport oxygen across the blood–brain barrier and increase the availability of oxygen to the metabolically very active neuronal tissues. Considering the phylogenetically basal position of the neuroglobins in the metazoan tree (Fig. 3), it is conceivable that oxygen-supplying nerve globins in fact emerged with the evolution of the neurosystem, and similar proteins may be found in other taxa as well. □

Methods

Cloning, sequencing, mapping and expression analysis

We used oligonucleotide primers (designed according to human ESTs R36571, AA324100 and F13069, and mouse EST AU036042) to amplify neuroglobin cDNAs from mouse and human brain RNA (accession numbers AJ245945 and AJ245946) by RT–PCR. Cloned cDNA was labelled²³ and used to detect the human *NGB* gene on PAC clone filters (http://www.rzpd.de). The *NGB* sequence (accession number AJ245944) was determined by GENEenterprise (Mainz). The mRNA start was identified by a 5' RACE kit (Life Technologies) and human total brain RNA (Research Genetics). Radiation hybrid mapping was performed using the GENEBRIDGE 4 RH panel (Research Genetics). An RNA Master Blot (Clontech) containing normalized amounts of poly(A)⁺ RNA from 50 human tissues was hybridized to a labelled²³ *NGB* cDNA probe. Signals were quantified on a Fuji BAS-1800 phosphorimager. For mRNA *in situ* hybridization, sagittal cryosections were prepared from adult Balb/C mouse brain²⁴. Forty-base oligonucleotide probes were 3' end labelled with digoxigenin (DIG)-dUTP (Roche). We carried out hybridization on frozen sections (fixed for 10 min in 4% paraformaldehyde/PBS) in 2× standard saline citrate (SSC: 0.3 M NaCl, 0.03 M Na-citrate)/50% formamide at 32°C. We detected label by alkaline phosphatase-coupled anti-DIG antibodies and nitroblue-tetrazolium salt/5-bromo-4-chloroindolylphosphate-toluidium salt (NBT/BIP) substrate.

Recombinant expression of neuroglobin

Mouse *Ngb* cDNA was cloned into pET-3a (ref. 25) and transformed into *Escherichia coli* BL21(DE3)pLys. We grew the bacteria at 25°C in LB medium containing 100 µg ml^{−1} ampicillin and 1 mM δ-aminolevulinic acid. Expression was induced by 1 mM IPTG. After 14–18 h, we collected the reddish bacteria, washed them with 25 mM Tris-HCl, pH 8, 10 mM EDTA and 0.9% glucose, and resuspended them in 50 mM Tris, 1 mM EDTA and 0.5 mM dithiothreitol supplemented with Roche Complete proteinase inhibitor mix. Bacteria were broken by freeze-thawing followed by ultrasonication. Cell debris was removed by centrifugation. Ngb was precipitated with 40–60% (NH₄)₂SO₄ and purified by size-exclusion chromatography (Sephacryl S-200, Pharmacia) followed by ion-exchange chromatography (Mono-Q, Pharmacia). The absorption spectrum was obtained from Ngb dissolved in 50 mM potassium phosphate, 1 mM EDTA, pH 7.4 at 25°C.

Oxygen-binding studies

We carried out oxygen-binding studies by measuring the absorption at 424 nm (de-oxy maximum) in a Gill cell. Because of partial irreversible oxygenation during purification, oxygen-binding curves were obtained with the bacterial supernatant after concentration by Centriscat C-4 filters (5K cut-off; Sartorius, Germany). Bacteria without Ngb expression did not show significant changes in absorption. For the analysis of native Ngb, we homogenized ten brains from adult Balb/C mice in ice-cold PBS (137 mM NaCl, 2.7 mM KCl, 7 mM Na₂HPO₄, 1.5 mM KH₂PO₄, pH 7.4) supplemented with 1 mM EDTA, 0.5 mM DTT and Roche Complete proteinase inhibitor mix. Proteins were precipitated with 35–65% (NH₄)₂SO₄ dissolved in potassium phosphate buffer and Ngb was partially purified by size-exclusion chromatography.

Phylogenetic inference

A multiple alignment of selected globin sequences was constructed with ClustalX²⁶ and corrected using a published alignment²⁷ and globin structural data. Phylogenetic analysis was carried out by the neighbour-joining method based on a PAM001 distance matrix²⁸. The reliability of trees was tested by bootstrap analysis with 100 replications.

Received 19 April; accepted 3 July 2000.

- Dickerson, R. E. & Geis, I. *Hemoglobin: Structure, function, evolution, and pathology* (Benjamin/Cummings, Menlo Park, California, 1983).
- Stryer, L. *Biochemistry* (Freeman, New York, 1995).
- Hardison, R. C. A brief history of hemoglobins: plant, animal, protist, and bacteria. *Proc. Natl. Acad. Sci. USA* **93**, 5675–5679 (1996).
- Gödecke, A. et al. Disruption of myoglobin in mice induces multiple compensatory mechanisms. *Proc. Natl. Acad. Sci. USA* **96**, 10495–10500 (1999).
- Boguski, M. S., Lowe, T. M. & Tolstoshev, C. M. dbEST—database for “expressed sequence tags”. *Nature Genet.* **4**, 332–333 (1993).
- Makalowski, W., Zhang, J. & Boguski, M. S. Comparative analysis of 1196 orthologous mouse and human full length mRNA and protein sequences. *Genome Res.* **6**, 846–857 (1996).
- Waterman, M. R. Spectral characterization of human hemoglobin and its derivatives. *Methods Enzymol.* **52**, 456–463 (1978).
- Strittmatter, P. & Burch, H. B. The heme protein in ganglia of *Spisula solidissima*. *Biochim. Biophys. Acta* **78**, 562–563 (1963).
- Dixon, B. & Pohajdak, B. Did the ancestral globin gene of plants and animals contain only two introns? *Trends Biochem. Sci.* **17**, 486–488 (1992).
- Go, M. Correlation of DNA exonic regions with protein structural units in haemoglobin. *Nature* **291**, 90–92 (1981).
- Hankeln, T., Friedl, H., Ebersberger, I., Martin, J. & Schmidt, E. R. A variable intron distribution in globin genes of *Chironomus*: evidence for recent intron gain. *Gene* **205**, 151–160 (1997).
- Logsdon, J. M., Stoltzfus, A. & Doolittle, W. F. Molecular evolution: recent cases of spliceosomal intron gain? *Curr. Biol.* **8**, R560–R563 (1998).
- Goodman, M., Moore, G. W. & Matsuda, G. Darwinian evolution in the genealogy of haemoglobin. *Nature* **253**, 603–608 (1975).
- Dewilde, S. et al. Globin and globin gene structure of the nerve myoglobin of *Aphrodite aculeata*. *J. Biol. Chem.* **271**, 19865–19870 (1996).
- Simons, P. C. & Satterlee, J. D. cDNA cloning and predicted amino acid sequence of *Glycera dibranchiata* monomer hemoglobin IV. *Biochemistry* **28**, 8525–8530 (1989).
- Zafar, R. S. et al. The cDNA sequences encoding two components of the polymeric fraction of the intracellular hemoglobin of *Glycera dibranchiata*. *J. Biol. Chem.* **265**, 21843–21851 (1990).
- Zafar, R. S., Chow, L. H., Stern, M. S., Vinogradov, S. N. & Walz, D. A. The heterogeneity of the polymeric intracellular hemoglobin of *Glycera dibranchiata* and the cDNA-derived amino acid sequence of one component. *Biochim. Biophys. Acta* **1041**, 117–122 (1990).
- Wittenberg, B. A., Brieht, R. W. & Wittenberg, J. B. Haemoglobins of invertebrate tissues - nerve haemoglobins of *Aphrodite*, *Aplysia*, and *Halosydna*. *Biochem. J.* **96**, 363–371 (1965).
- Wittenberg, J. B. Functions of cytoplasmic hemoglobins and myohemerythrin. *Adv. Comp. Environ. Physiol.* **13**, 60–85 (1992).
- Vandergon, T. L., Riggs, C. K., Gorr, T. A., Colacino, J. M. & Riggs, A. F. The mini-hemoglobins in neural and body wall tissue of the nemertean worm, *Cerebratulus lacteus*. *J. Biol. Chem.* **273**, 16998–17011 (1998).
- Kraus, D. W. & Colacino, J. M. Extended oxygen delivery from the nerve hemoglobin of *Tellina alternata* (Bivalvia). *Science* **232**, 90–92 (1986).
- Schmidt, R. F. & Thews, G. *Physiologie des Menschen* (Springer, Berlin, Heidelberg, New York, 1997).
- Feinberg, A. & Vogelstein, B. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **132**, 6–13 (1983).
- Rentrop, M., Knapp, B., Winter, H. & Schweizer, J. Aminoacylsilane-coated glass slides as support for *in situ* hybridisation of keratin cDNAs to frozen tissue sections under varying fixation and pretreatment conditions. *Histochem. J.* **18**, 271–276 (1986).
- Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.* **185**, 60–89 (1990).

26. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**, 4876–82 (1997).
27. Burmester, T. & Hankeln, T. A globin gene of *Drosophila melanogaster*. *Mol. Biol. Evol.* **16**, 1809–1811 (1999).
28. Felsenstein, J. PHYLIP (Phylogeny Inference Package) Version 3.5c (Dept Genetics, Univ. Washington, Seattle, 1993).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We wish to thank R. Gebhardt and N. Hellmann for their help with the oxygen-binding studies; G. Ungerechts for his assistance in the cloning; E. Jaenicke for running the FPLC; L. Moens and S. Dewilde for sharing experimental protocols and for discussions; and H. Decker, E. R. Schmidt and J. Markl for excellent working facilities, continuous support and valuable suggestions. This work is supported by the Deutsche Forschungsgemeinschaft (DFG) and the Naturwissenschaftlich-Medizinisches Forschungszentrum (NMFZ) Mainz.

Correspondence and requests for material should be addressed to T.B. (e-mail: burmeste@mail.uni-mainz.de) or T.H. (e-mail: hankeln@molgen.biologie.uni-mainz.de).

Immune control of HIV-1 after early treatment of acute infection

Eric S. Rosenberg, Marcus Altfeld, Samuel H. Poon, Mary N. Phillips, Barbara M. Wilkes, Robert L. Eldridge, Gregory K. Robbins, Richard T. D'Aquila, Philip J. R. Goulder & Bruce D. Walker

Partners AIDS Research Center and Infectious Disease Division, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts 02114, USA

Virus-specific T-helper cells are considered critical for the control of chronic viral infections^{1,2}. Successful treatment of acute HIV-1 infection leads to augmentation of these responses^{3–5}, but whether this enhances immune control has not been determined. We administered one or two supervised treatment interruptions to eight subjects with treated acute infection, with the plan to restart therapy if viral load exceeded 5,000 copies of HIV-1 RNA per

millilitre of plasma (the level at which therapy has been typically recommended) for three consecutive weeks, or 50,000 RNA copies per ml at one time. Here we show that, despite rebound in viraemia, all subjects were able to achieve at least a transient steady state off therapy with viral load below 5,000 RNA copies per ml. At present, five out of eight subjects remain off therapy with viral loads of less than 500 RNA copies per ml plasma after a median 6.5 months (range 5–8.7 months). We observed increased virus-specific cytotoxic T lymphocytes and maintained T-helper-cell responses in all. Our data indicate that functional immune responses can be augmented in a chronic viral infection, and provide rationale for immunotherapy in HIV-1 infection.

Many viruses, including Epstein–Barr virus and cytomegalovirus, are not cleared by the host but are controlled by an effective immune response⁶. Others, including hepatitis C virus and HIV-1, are usually characterized by lack of immune control and progressive infection. This has been linked to a lack of virus-specific T-helper-cell responses, which are required for maintenance of effective cytotoxic T-lymphocyte (CTL) function during the chronic phase of infection^{1–5,7–11}. In contrast, immediate treatment of acute HIV-1 infection with highly active antiretroviral therapy (HAART) leads to generation of strong HIV-1 Gag-specific T-helper-cell responses³, and anecdotal cases of control of viraemia after treatment of acute HIV infection^{4,12,13} have been associated with these responses^{4,12}. We therefore examined the antiviral effectiveness of cellular immune responses in individuals effectively treated with HAART during acute or early HIV-1 infection.

Sixteen individuals were identified with symptomatic acute HIV-1 infection¹⁴, as were two subjects with recent infection (<180 d). All were started on HAART, most within 72 h of diagnosis (range 1–34 d). Gag-specific T-helper-cell responses became detectable in the 15 subjects in whom treatment led to complete suppression (<50 RNA copies per ml) of plasma viraemia, and were maintained at levels similar to those in subjects who spontaneously control viraemia (Fig. 1). Untreated subjects with documented acute infection¹⁵ had weak or absent Gag-specific T-helper-cell responses ($P < 0.01$). Similarly, T-helper-cell responses were lower in subjects started on therapy during chronic infection ($P < 0.01$), with the only clearly positive response in a subject who initiated therapy despite only 4,000 HIV-1 RNA copies per ml plasma. HIV-1-specific CTL responses at 1 yr, measured by interferon (IFN)- γ enzyme-linked immunospot (Elispot) assay, were modest in the 18 individuals with treated primary HIV-1 infection, with a cumulative mean

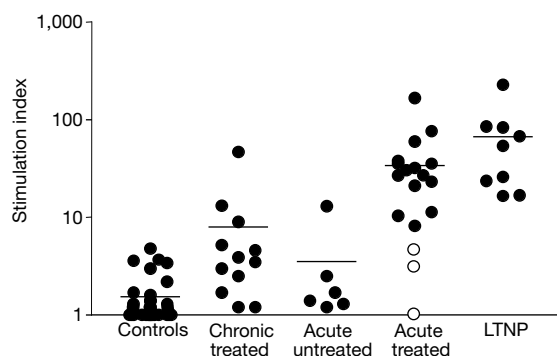


Figure 1 HIV-1 Gag-specific T-helper-cell responses in HIV-1 infection. Values at 1 yr of treated acute infection (acute treated) were compared with uninfected subjects (control), subjects first treated in the chronic phase of infection (chronic treated), subjects with untreated acute infection (acute untreated), and subjects with long-term non-progressing HIV-1 infection who are able to control viraemia to less than 1,000 RNA copies per ml plasma in the absence of therapy (LTNP). Lymphocyte proliferation to Gag protein is shown as stimulation index. Elispot assays showed that the IFN- γ production was specifically induced, consistent with a Th-1 type immune response (data not shown). Open symbols indicate subjects with treated acute infection infected with drug resistant HIV.

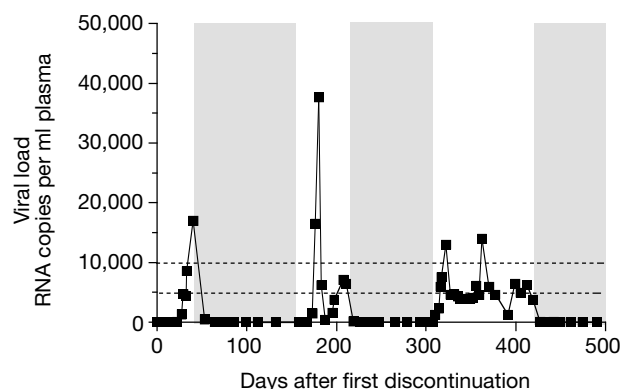


Figure 2 Viral load in a pilot study of successive treatment interruptions in a 32-year-old male with treated acute HIV-1 infection. This subject was treated with HAART (ZDV, 3TC and IDV) for 19 months before interruption in therapy. Time zero represents the first therapy interruption. Shaded areas represent treatment with the same HAART regimen. The second therapy interruption was off protocol owing to viral hepatitis. Cumulative HIV-1-specific CTL responses, as measured by Elispot, increased from 2,360 SFCs per 10^6 PBMCs before the first interruption to a peak of 5,760 SFCs per 10^6 PBMCs after the last treatment interruption (data not shown).

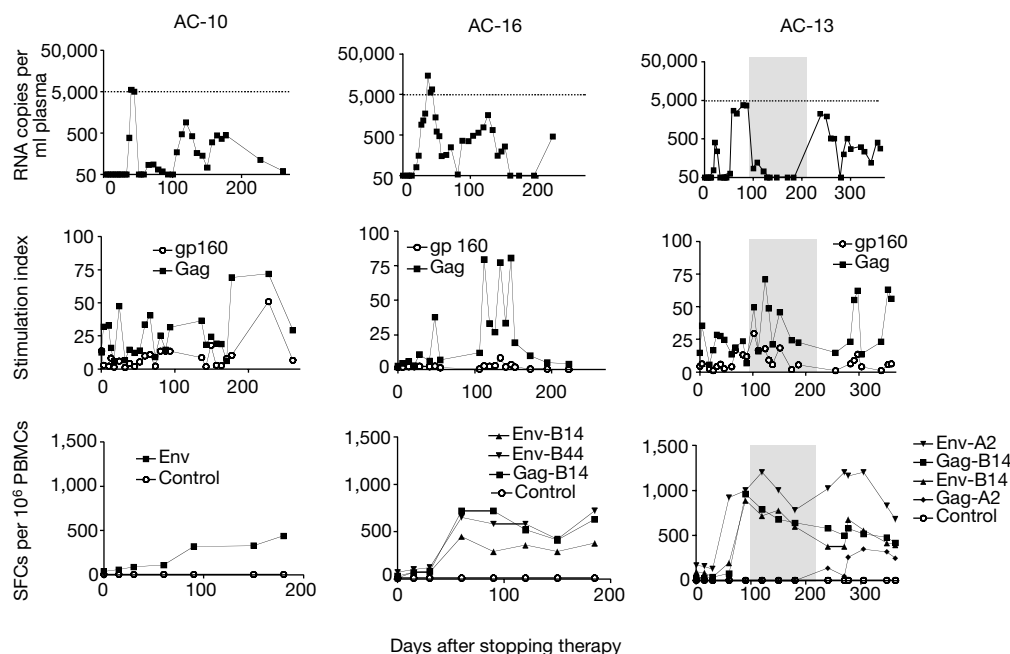


Figure 3 Virology and immunology in three subjects who controlled viraemia after the first treatment interruption. In all three, plasma virus load rose but then dropped, remaining consistently below 5,000 copies of HIV-1 RNA per ml (upper row). Gag-specific T-helper-cell responses were maintained in all three subjects (middle row). CTL responses increased significantly from baseline values and are shown as responses against

individual CTL epitopes with HLA restriction (lower row). CTL responses were confirmed by cloning and testing in cytolytic assays, as described²⁹. The shaded area indicates antiviral therapy re-initiated by subject 3, despite not meeting protocol indication for re-treatment.

of 460 spot-forming cells (SFCs) per 10^6 peripheral blood mononuclear cells (PBMCs) (s.d. 650 SFCs per 10^6 PBMCs) directed against only a few CTL epitopes (median 2, range 0–7).

To determine whether the immune responses induced with early treatment of acute infection were functional, we obtained Institutional Review Board approval for a pilot study to discontinue antiviral treatment. Therapy resumption was required for any viral load $\geq 10,000$ HIV-1 RNA copies per ml. Discontinuation in the initial subject AC-01 led to recurrence of viraemia, which was controlled with re-institution of therapy (Fig. 2). Subsequent acute hepatitis A virus infection necessitated an unplanned treatment interruption owing to liver function abnormalities, at which time virus rebounded more rapidly and to a higher peak, but then declined without treatment to 490 RNA copies per ml plasma over the next 6 d. Therapy was subsequently restarted after liver enzymes had normalized. Approval was then granted to discontinue therapy again and resume for a viral load $> 50,000$ copies per ml. Viraemia was detected earlier (9 d), rose to a lower peak and declined to 4,680 copies 5 d later (Fig. 2), which is below the level at which therapy has been typically recommended¹⁶. A subsequent episode of bacterial pharyngitis was associated with a modest increase in viraemia, which declined without the re-institution of therapy. CD4⁺ T-cell counts remained stable throughout (data not shown). Although not required by protocol, the patient decided to resume treatment when viral load was 3,820 RNA copies per ml.

On the basis of these results, we enrolled eight new subjects in a revised protocol, mandating the re-initiation of therapy for a viral load $> 5,000$ RNA copies per ml for three consecutive weeks or a single value $\geq 50,000$ copies. These eight (Table 1) were indistinguishable in terms of viral load, CD4 counts and HIV-1-specific T-helper and CTL responses from the other eligible subjects who did not interrupt therapy.

During the first treatment interruption, plasma virus became detectable in all subjects after a median of 17 d (range 7–38 d). Although early peaks in viraemia were observed, viral load dropped to consistently below 5,000 RNA copies per ml plasma in three

subjects (Fig. 3). Two remain off therapy after 8.7 and 7.4 months, respectively. One (AC-13) elected to restart HAART after 91 d, having never exceeded 5,000 RNA copies per ml plasma. He subsequently discontinued therapy again and has a viral load of 280 RNA copies per ml plasma 5.4 months later. In these three subjects, Gag-specific T-helper-cell responses fluctuated but were maintained during and after treatment interruption (Fig. 3). Virus-specific CTL responses (Fig. 3) were meagre but detectable at the time of therapy interruption, and directed against a median of three CTL epitopes (range 1–3). Exposure to virus in these three subjects was associated with greater than tenfold increases in median CTL from 110 SFCs per 10^6 PBMCs (range 40–300) to 1,730 SFCs per 10^6 PBMCs (range 440–2,850) ($P < 0.001$). CTL responses also broadened to target more epitopes in one of the three subjects (AC-13).

Five of the eight individuals were required to restart therapy after the first treatment interruption, four after plasma virus load exceeded 50,000 HIV-1 RNA copies per ml plasma and one because the viral load exceeded 5,000 RNA copies per ml plasma for three consecutive weeks (Fig. 4a). All five underwent a second treatment interruption. Plasma viraemia rose to a lower peak level in all ($P < 0.01$), and then dropped to less than 5,000 RNA copies per ml without treatment (Fig. 4b). Only one subject (AC-04) met criteria

Table 1 Subjects with acute or early HIV infection in a prospective trial of treatment interruption

Subject	Age	Therapy*	Days from presentation to initiation of therapy	Days on therapy before interruption
AC-02	42	ZDV,3TC,NFV	3	680
AC-04	35	ZDV,3TC,NFV	7	594
AC-05	37	ZDV,3TC,NFV	2	653
AC-06	36	D4T,3TC,NFV	4	546
AC-10†	30	D4T,3TC,IND	4	573
AC-13	32	D4T,3TC,IND	15	358
AC-14†	44	D4T,3TC,IND	34	383
AC-16	34	ZDV,3TC,IND	2	1081

* ZDV, zidovudine; D4T, stavudine; 3TC, lamivudine; IND, indinavir; NFV, nelfinavir.

† ELISA positive but infection within 180 days at time therapy initiated.

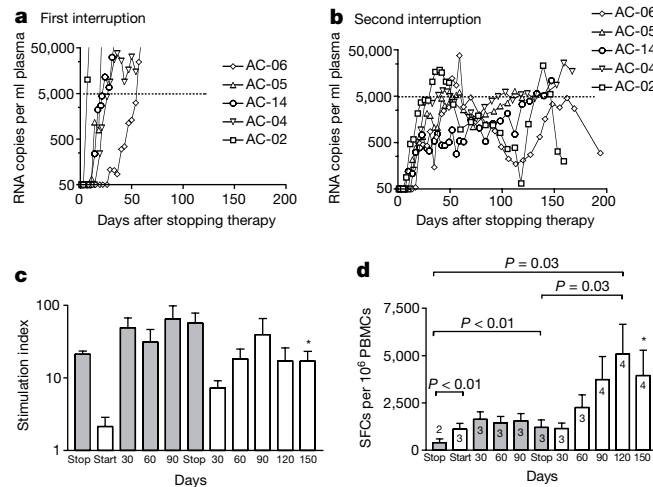


Figure 4 Virology and immunology in subjects requiring repeated treatment interruption. **a**, Five subjects met viraemia criteria to re-institute therapy during the first treatment interruption. Viral load (RNA copies per ml plasma) at the time of therapy resumption was as follows: AC-02, 187,000; AC-04, 14,800; AC-05, 116,000; AC-06, 139,000; AC-14, 111,000. Upon re-initiation of therapy, HIV-1 load declined below limits of detection in all individuals within 35–130 d (data not shown). **b**, After a median of 16 weeks of additional therapy (range 13–30 weeks), therapy was again interrupted. **c**, **d**, Gag-specific T-helper-cell responses (**c**) and HIV-1-specific CTL responses (**d**) were assessed at the time therapy was stopped (stop), at the time of protocol-mandated re-institution of therapy

(start), and at designated days in between. Shaded bars indicate assays performed when subjects were on HAART. CTL magnitudes are given as mean of total CTL responses per individual, calculated as the sum of responses against optimal CTL epitopes. The median number of CTL epitopes recognized per individual is indicated within the bars, and reached a median of four per individual after the last interruption (range 2–9). CTL responses were confirmed by cloning and testing in cytolytic assays. *P* values are calculated for changes in CTL magnitude. Asterisk indicates that analysis was limited to four subjects.

to restart therapy when viral load reached 17,100 RNA copies per ml plasma after 5.5 months of treatment interruption. Two remain off therapy after 6.5 and 5.3 months, respectively, with most recent plasma viral loads of 290 and 200 RNA copies per ml, respectively. Two subjects elected to restart therapy despite not having met protocol criteria, one with a viral load of 4,600 RNA copies per ml plasma after 4 months off therapy, and one with a viral load trending up at 10,860 RNA copies per ml plasma after 5 months.

In these five individuals, Gag-specific T-helper-cell responses declined during the first treatment discontinuation but rose to levels higher than baseline with the re-initiation of therapy (Fig. 4c). During the second interruption in treatment, the stimulation index transiently declined and then rose in the absence of therapy and despite maintained low levels of plasma virus. In addition, the mean length of time off therapy was extended from 38 d to 157 d. Virus-specific CTL responses increased significantly after re-exposure to viral antigen during the first treatment interruption ($P < 0.01$), and had broadened to a median of three epitopes targeted (Fig. 4d). Two subjects who previously had no detectable CTL responses (AC-14 and AC-04) developed these for the first time after stopping therapy. CTL responses persisted at high levels after re-institution of HAART, despite undetectable viral RNA in plasma. During the second treatment interruption, CTL again increased (Fig. 4d) and more epitopes were targeted ($P < 0.01$, compared with baseline before the first treatment interruption). Dynamic immunologic control was further suggested by subject AC-02, who experienced a late increase in viral load to almost 20,000 RNA copies per ml plasma that spontaneously declined to 160 copies per ml. The decrease in viraemia was associated with further broadening of the CTL response (data not shown). In contrast, subjects who did not interrupt treatment had no increase in magnitude or breadth of HIV-1-specific CTL responses (data not shown).

At the time of writing, five of the eight individuals remain off therapy an average of 2.7 yr after initial infection, with most recent viral loads less than 500 RNA copies per ml plasma (Figs 3 and 4b).

This is highly significant ($P < 0.001$) compared with historical controls from the Multicenter AIDS Cohort Study (MACS), in which only 4 out of 109 subjects with documented acute infection had viral loads less than 500 RNA copies per ml plasma 30 months after infection (ref. 17; and A. Munoz, personal communication). Comparison of those with viral loads less than 5,000 copies per ml was also highly significant, with 5 out of 8 in our cohort compared with 16 out of 109 in the MACS after a comparable period of infection ($P < 0.001$). None of the infected subjects recognized all predicted viral CTL epitopes for their expressed human leukocyte antigen (HLA) class I alleles, suggesting that further broadening of the immune response under HAART treatment may confer additional benefit on immune control^{18–20}. This is a currently testable hypothesis, for example by using dendritic cells pulsed with synthetic viral peptides containing CTL and T-helper-cell epitopes²¹, or by therapeutic immunization. Such approaches might decrease the risks inherent in treatment interruptions, particularly the potential development of antiviral drug resistance.

Our data show that functional HIV-1-specific CD4 and CD8 T-cell responses can be enhanced with early treatment of acute infection. They contrast the reported lack of immune control after treatment interruption in chronic infection²². Our data do not allow one to assess the contribution of sequential treatment interruptions to immunologic control of viraemia, as we re-instituted therapy before viral set point was achieved in the five subjects who underwent two interruptions in therapy. Whether effective immune augmentation can be achieved in chronic infection through repeated structured treatment interruptions is an important issue that is yet to be addressed, as is the durability of the observed immune control in our cohort. The effect on clinical outcome will require further longitudinal studies. Viral load has been used as a surrogate for efficacy in drug trials in HIV-1 infection, and if the same criteria are applied, one would predict a survival advantage. The successful containment of viraemia in most of these subjects provides rationale to support the practice of treating individuals with HAART during acute or early HIV-1 infection, and provides

rationale to explore immunotherapeutic interventions in both acute and chronic HIV-1 infection. □

Methods

Subjects

Sixteen individuals were identified with symptomatic acute HIV-1 infection¹⁴, defined by the presence of HIV-1 RNA in the plasma, and negative or weakly positive antibody tests (HIV 1/2 enzyme-linked immunosorbent assay (ELISA) and HIV-1 western blot). We identified two additional subjects with a recent syndrome compatible with acute HIV-1 infection and positive antibodies to HIV-1, in whom an ELISA-based assay confirmed HIV-1 infection within the previous 180 d (ref. 23). The median plasma HIV-1 load was 4.85×10^6 RNA copies per ml plasma (range $0.25\text{--}30.7 \times 10^6$ copies per ml) in the subjects diagnosed before seroconversion. The median CD4⁺ T-cell count before initiation of therapy was 365 cells per mm³ (range $130\text{--}1,020$ cells per mm³; normal $380\text{--}1,630$ cells per mm³). The two subjects identified after HIV-1 seroconversion had viral loads of 0.4×10^6 and 0.95×10^6 copies per ml and CD4⁺ T-cell counts of 920 and 980 cells per mm³, respectively. In addition, 6 subjects identified within 6–12 months of documented acute HIV-1 infection were used as an untreated control group. All individuals initiated HAART at the time of diagnosis with anti-retroviral regimens comprising either two nucleoside reverse transcriptase inhibitors combined with a protease inhibitor or a non-nucleoside reverse transcriptase inhibitor, or both. Additional groups of study subjects included 6 subjects identified within 6–12 months of documented HIV-1 seroconversion, and evaluated before the initiation of therapy¹⁵; 12 subjects with chronic HIV-1 infection whose virus had been effectively suppressed by HAART (viral loads < 50 RNA copies per ml plasma) for at least 12 months, 7 of whom had CD4⁺ T-cell counts > 500 cells per mm³ before initiating therapy; and 9 subjects with chronic HIV-1 infection who controlled viral replication in the absence of anti-retroviral therapy (viral load < 1,000 RNA copies per ml, CD4⁺ T cells > 500 per mm³, > 10 yr infection).

Treatment interruption

Entry criteria included treatment with HAART before or around the time of HIV-1 seroconversion, consistent suppression of viraemia to below levels of quantification (< 50 RNA copies per ml plasma) for at least 8 months, and lack of evidence of significant mutations conferring drug resistance²⁴. HIV-1-specific T-helper-cell function before discontinuation had to exceed a stimulation index of 10, and net counts per minute (c.p.m.) had to be ≥ 800 . All anti-retroviral medications were discontinued simultaneously.

HIV-1-specific T-helper-cell assays

We carried out lymphocyte proliferation assays with baculovirus-derived HIV-1 Gag and gp160 protein. PBMCs were incubated with peptide ($5 \mu\text{g ml}^{-1}$) for 6 d and then pulsed with [³H]thymidine at $1.0 \mu\text{Ci}$ per well for 6 h, as described³. For the purposes of data interpretation, a stimulation index of 5 or greater was considered significant.

IFN- γ Elispot assays of HIV-1-specific CD4 and CD8 T-cell responses

Synthetic peptides for detection of CTL responses comprised 259 overlapping peptides of 15–20 amino acids, overlapping by 10–11 amino acids, spanning the entire p17 Gag, p24 Gag, Nef, gp41, gp120 and reverse transcriptase B-clade coding regions, and a panel of described optimal CTL epitopes for the HLA-type of the individual tested²⁵. A median of 16 (range 5–28) described optimal CTL epitopes were tested per individual. For CTL detection by Elispot, we plated fresh or frozen PBMCs in 96-well polyvinylidene difluoride backed plates (MAIP S45, Millipore) that had been previously coated with $0.5 \mu\text{g ml}^{-1}$ of an anti-IFN- γ monoclonal antibody 1-D1k (Mabtech, Stockholm, Sweden) overnight at 4 °C. We added PBMCs to the wells at 50,000 to 100,000 cells per well, and processed and counted plates as described²⁶. We counted IFN- γ -producing cells by direct visualization and express them as spot-forming cells (SFCs). The negative controls were subtracted, and in all cases were less than 40 SFCs per 10^6 input cells. We generated CTL clones to confirm the cytolytic ability of the antigen-specific responses, as described²⁷. We quantitated CD4⁺ T-helper-cell responses using a similar assay, with minor differences. We incubated PBMCs (1×10^6) overnight in a 48-well plate with HIV-1 p24 antigen, baculovirus control protein or medium alone. Cells were then transferred to Elispot plates pre-coated with $2.5 \mu\text{g ml}^{-1}$ monoclonal IFN- γ antibody (Endogen, Woburn, MA), and incubated for 24–48 h. Plates were developed and counted as above. For each response to an overlapping peptide, the optimal CTL epitope contained within the 15–20 mer was characterized and used for the subsequent quantification of CTL responses.

Sequencing of HIV-1 protease and reverse transcriptase

Viron RNA was extracted (QiAmp Viral RNA kit, Qiagen) from centrifuged plasma before complementary DNA synthesis (Superscript II, Gibco-BRL). Two rounds of polymerase chain reaction were done with a nested set of primers amplifying a fragment from p6 Gag through codon 350 in the reverse transcriptase (XL rTth DNA PCR kit, PE Biosystems), and then sequenced directly and analysed as described²⁸.

Statistical analysis

We carried out statistical analysis using GraphPad InStat version 3.00 for Windows 95 (GraphPad Software, San Diego, CA). Statistical significance (*P* values) of results were calculated by two-tailed *t*-test and by χ^2 analysis.

Received 17 July; accepted 31 August 2000.

- Matloubian, M., Concepcion, R. J. & Ahmed, R. CD4⁺ T cells are required to sustain CD8⁺ cytotoxic T-cell responses during chronic viral infection. *J. Virol.* **68**, 8056–8063 (1994).
- Battagay, M. et al. Enhanced establishment of a virus carrier state in adult CD4⁺ T-cell-deficient mice. *J. Virol.* **68**, 4700–4704 (1994).
- Rosenberg, E. S. et al. Vigorous HIV-1-specific CD4⁺ T cell responses associated with control of viremia. *Science* **278**, 1447–1450 (1997).
- Oxenius, A. et al. Early highly active antiretroviral therapy for acute HIV-1 infection preserves immune function of CD8⁺ and CD4⁺ T lymphocytes. *Proc. Natl Acad. Sci. USA* **97**, 3382–3387 (2000).
- Malhotra, U. et al. Effect of combination antiretroviral therapy on T-cell immunity in acute human immunodeficiency virus type 1 infection. *J. Infect. Dis.* **181**, 121–131 (2000).
- Kalams, S. A. & Walker, B. D. The critical need for CD4 help in maintaining effective cytotoxic T lymphocyte responses. *J. Exp. Med.* **188**, 2199–2204 (1998).
- von Herrath, M. G., Yokoyama, M., Dockter, J., Oldstone, M. B. & Whitton, J. L. CD4-deficient mice have reduced levels of memory cytotoxic T lymphocytes after immunization and show diminished resistance to subsequent virus challenge. *J. Virol.* **70**, 1072–1079 (1996).
- Zajac, A. J. et al. Viral immune evasion due to persistence of activated T cells without effector function. *J. Exp. Med.* **188**, 2205–2213 (1998).
- Kalams, S. A. et al. Association between virus-specific cytotoxic T-lymphocyte and helper responses in human immunodeficiency virus type 1 infection. *J. Virol.* **73**, 6715–6720 (1999).
- Gerlach, J. T. et al. Recurrence of hepatitis C virus after loss of virus-specific CD4(+) T-cell response in acute hepatitis C. *Gastroenterology* **117**, 933–941 (1999).
- Lechner, F. et al. Analysis of successful immune responses in persons infected with hepatitis C virus. *J. Exp. Med.* **191**, 1499–1512 (2000).
- Liszewicz, J. et al. Control of HIV despite the discontinuation of antiretroviral therapy. *N. Engl. J. Med.* **340**, 1683–1684 (1999).
- Ortiz, G. M. et al. HIV-1-specific immune responses in subjects who temporarily contain virus replication after discontinuation of highly active antiretroviral therapy. *J. Clin. Invest.* **104**, R13–R18 (1999).
- Kahn, J. O. & Walker, B. D. Acute human immunodeficiency virus type 1 infection. *N. Engl. J. Med.* **339**, 33–39 (1998).
- Lori, F. et al. Treatment of human immunodeficiency virus infection with hydroxyurea, didanosine, and a protease inhibitor before seroconversion is associated with normalized immune parameters and limited viral reservoir. *J. Infect. Dis.* **180**, 1827–1832 (1999).
- Carpenter, C. C. et al. Antiretroviral therapy for HIV infection in 1997. Updated recommendations of the International AIDS Society/USA panel. *J. Am. Med. Assoc.* **277**, 1962–1969 (1997).
- Lyles, R. H. et al. Natural history of human immunodeficiency virus type 1 viremia after seroconversion and proximal to AIDS in a large cohort of homosexual men. Multicenter AIDS Cohort Study. *J. Infect. Dis.* **181**, 872–880 (2000).
- Pantaleo, G. et al. Major expansion of CD8⁺ T cells with a predominant V beta usage during the primary immune response to HIV. *Nature* **370**, 463–467 (1994).
- Autran, B. et al. Positive effects of combined antiretroviral therapy on CD4⁺ T cell homeostasis and function in advanced HIV disease. *Science* **277**, 112–116 (1997).
- Evans, D. T. et al. Virus-specific cytotoxic T-lymphocyte responses select for amino-acid variation in simian immunodeficiency virus Env and Nef. *Nature Med.* **5**, 1270–1276 (1999).
- Dhodapkar, M. V. et al. Rapid generation of broad T-cell immunity in humans after a single injection of mature dendritic cells. *J. Clin. Invest.* **104**, 173–180 (1999).
- Davey, R. T. Jr et al. HIV-1 and T cell dynamics after interruption of highly active antiretroviral therapy (HAART) in patients with a history of sustained viral suppression. *Proc. Natl Acad. Sci. USA* **96**, 15109–15114 (1999).
- Janssen, R. S. et al. New testing strategy to detect early HIV-1 infection for use in incidence estimates and for clinical and prevention purposes. *J. Am. Med. Assoc.* **280**, 42–48 (1998).
- Little, S. J. et al. Reduced antiretroviral drug susceptibility among patients with primary HIV infection. *J. Am. Med. Assoc.* **282**, 1142–1149 (1999).
- Brander, C. & Walker, B. D. in *Molecular Immunology Data Base* (eds Korber, B. T. et al.) IV-50–IV-60 (Los Alamos National Laboratory, Los Alamos, New Mexico, 1997).
- Goulder, P. J. et al. Differential narrow focusing of immunodominant human immunodeficiency virus gag-specific cytotoxic T-lymphocyte responses in infected African and caucasoid adults and children. *J. Virol.* **74**, 5679–5690 (2000).
- Walker, B. D. et al. Long-term culture and fine specificity of human cytotoxic T-lymphocyte clones reactive with human immunodeficiency virus type 1. *Proc. Natl Acad. Sci. USA* **86**, 9514–9518 (1989).
- Martinez-Picado, J., Sutton, L., De Pasquale, M. P., Savara, A. V. & D'Aquila, R. T. Human immunodeficiency virus type 1 cloning vectors for antiretroviral resistance testing. *J. Clin. Microbiol.* **37**, 2943–2951 (1999).
- Kalams, S. A. et al. Levels of human immunodeficiency virus type 1-specific cytotoxic T-lymphocyte effector and memory responses decline after suppression of viremia with highly active antiretroviral therapy. *J. Virol.* **73**, 6721–6728 (1999).

Acknowledgements

This work was supported by the Doris Duke Charitable Foundation (B.D.W., E.S.R.), the NIH (E.S.R., R.D., B.D.W.), the Elizabeth Glaser Pediatric AIDS Foundation (P.J.R.G.), the Deutscher Akademischer Austauschdienst (M.A.), the UK Medical Research Council (P.J.R.G.) and several private donors. We thank A. Munoz for conducting the analysis comparing our data with the MACS data. We also thank the subjects who participated in these trials, M. S. Hirsch and D. T. Scadden for comments on the manuscript; B. Dale for supplying kits for viral load testing; P. Sax and B. Davis for help in recruiting study subjects; M. J. Ferraro, J. Eversley and S. Masci, for technical support; and B. Rawal for performing modified HIV-1 ELISA assays. P.J.R.G. is an Elizabeth Glaser Scientist and B.D.W. is a Doris Duke Distinguished Clinical Science Professor.

Correspondence and requests for materials should be addressed to B.D.W. (e-mail: bwalker@helix.mgh.harvard.edu).

arrow encodes an LDL-receptor-related protein essential for Wingless signalling

Marcel Wehrli*†‡, Scott T. Dougan§, Kim Caldwell§, Louise O'Keefe§, Stephanie Schwartz*§, Dalit Vaizel-Ohayon||, Eyal Schejter||, Andrew Tomlinson† & Stephen DiNardo*§

* University of Pennsylvania School of Medicine, BRB II Room 1223, 421 Curie Boulevard, Philadelphia, Pennsylvania 19104, USA

† Columbia University, 710 West 168th Street, New York, New York 10032, USA

§ Rockefeller University, 1230 York Avenue, New York, New York, 10021, USA

|| Weizmann Institute of Science, Rehovot 76100, Israel

‡ These authors contributed equally to this work

The Wnt family of secreted molecules functions in cell-fate determination and morphogenesis during development in both vertebrates and invertebrates (reviewed in ref. 1). *Drosophila* Wingless is a founding member of this family, and many components of its signal transduction cascade have been identified, including the Frizzled class of receptor. But the mechanism by which the Wingless signal is received and transduced across the membrane is not completely understood. Here we describe a gene that is necessary for all Wingless signalling events in *Drosophila*. We show that *arrow* gene function is essential in cells receiving Wingless input and that it acts upstream of Dishevelled. *arrow* encodes a single-pass transmembrane protein, indicating that it may be part of a receptor complex with Frizzled class proteins. Arrow is a low-density lipoprotein (LDL)-receptor-related protein (LRP), strikingly homologous to murine and human LRP5 and LRP6. Thus, our data suggests a new and conserved function for this LRP subfamily in Wingless/Wnt signal reception.

In each embryonic segment, Wingless (Wg) is expressed anterior to cells expressing Engrailed (En) which co-express the signal Hedgehog (Hh). Most of the pattern across each segmental field is organized by these two signals. For example, Wg has two roles in

patterning the segment. First, between 3 and 6 hours after egg laying (a.e.l.), early Wg signalling is necessary for the continued expression of En and Hh in the adjacent cell row. This ensures the maintenance of the parasegmental subdivisions of the body axis, generating the segmental body plan observed in larval cuticle preparations (Fig. 1a). Between 6 and 9 hours a.e.l., Wg signalling assigns specific cell fates. Within each segment, ventral epidermal cells secrete either smooth cuticle or protrusions called denticles. The double row of En/Hh-expressing cells straddles the boundary between smooth cuticle and the first denticle row (Fig. 1b). Wg signalling instructs anterior En cells to adopt smooth cell fate². When Wg function is eliminated after 8 hours a.e.l., the anterior En cell incorrectly adopts a denticle fate (Fig. 1c). Similarly, anterior En cells adopt denticle fates in *arrow* mutants (Fig. 1d). These embryos have wild-type segmentation because early Wg signalling proceeds normally in zygotic *arrow* mutants. However, when we removed both maternal and zygotic *arrow* function (Fig. 1e; hereafter referred to as *arr^{null}* embryos), the resulting embryos were indistinguishable from *wg* null mutants. As the maintenance of En expression depends on the reception of the early Wg signal, we assayed *arr^{null}* embryos for En maintenance. En expression is initiated properly in *wg* and *arr^{null}* embryos (Fig. 1f, h, j) but then fades (Fig. 1g, i, k). In addition, Wg protein is distributed normally in *arr^{null}* mutant embryos, as compared with wild-type embryos (Fig. 1l; and data not shown). In particular, in *arr^{null}* embryos Wg is still detectable within cells distant from the Wg-expressing cells. Thus, *arr* mutations do not affect the production of ligand, nor its distribution across the epithelium, but nevertheless block the two Wg-dependent patterning events in embryonic epidermis. Furthermore, *arrow* function is required for all other Wg signalling events that we have tested, such as in embryonic midgut (data not shown), as well as in all facets of imaginal disk patterning (see below). Thus, *arrow* conforms to the expectation for a gene encoding an essential component of the Wg pathway.

We cloned the *arrow* gene using a lethal P-element insertion (1(2)k08131) that fails to complement *arrow*. The longest complementary DNA (6,012 base pairs (bp); GenBank AY005815, CT15575) restores smooth cuticle to alternate segments of *arr^{null}* embryos when expressed under control of Prd-GAL4, indicating rescue of Arrow function and Wg signalling (Fig. 2a, compare with Fig. 1e).

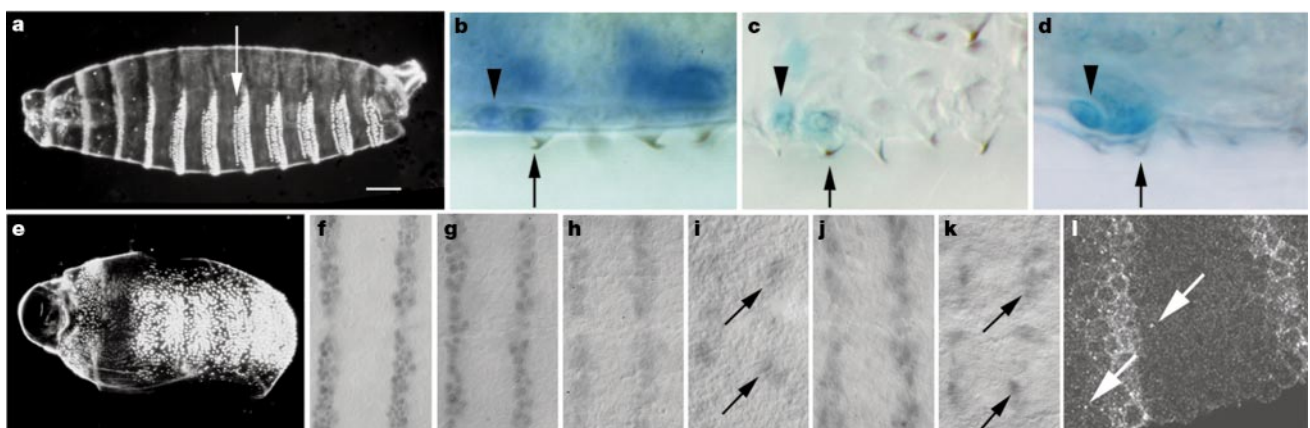


Figure 1 Arrow is required for Wg-dependent cell fates. Shown are larval cuticles, anterior left, ventrolateral (a), ventral (e–k) and optical sections (b–d). a, Wild type (WT); normal segmentation and patterning generates regions of alternating denticle and smooth cuticle. Arrow indicates first denticle row of abdominal segment four. b, WT; first four rows of a single denticle belt. Row 1 (arrow) is formed by the posterior of the two overlying EnlacZ-expressing cells (blue; b–d); the anterior En cell (arrowhead) secretes smooth cuticle. Staining above hypodermis is irrelevant CNS expression. c, *wg⁸*, inactivated at 8 hours a.e.l.; the anterior En cell (arrowhead) now produces a denticle. d, Zygotic *arrow* mutant; the anterior En cell (arrowhead) now adopts denticle fate. Ectopic denticles are also found just posterior to the belt³⁰. e, *arr^{null}* embryo; no smooth cuticle is produced, as

in *wg* null mutants (not shown). f–k, Anti-En monoclonal antibody 4D9 staining of one and a half parasegments of filleted epidermis. f, WT, 3.5 hours a.e.l.; En expression is initiated in narrow stripes. g, WT, 6 hours a.e.l.; En is maintained. h, i, *wg* null; En expression is induced normally (h) but fades from the epidermis (i), although it remains in underlying neuroblasts (arrows). j, k, *arr^{B/Df(2R)AA1}* maternal and zygotically mutant embryos; En expression is induced normally (j), but fades from the epidermis (k), although it remains in underlying neuroblasts (arrows). l, Anti-Wg, monoclonal antibody 4D4. *arr^{null}*, ~4.5 hours a.e.l., one parasegment magnified to visualize Wg protein (arrows) distant from Wg-producing cells. Scale bars, 70 μm (a); 6 μm (b–d); 50 μm (e); 25 μm (f–k); 9 μm (l).

Initially, *arrow* messenger RNA is distributed uniformly in the embryo, owing to maternal contribution, but by stage 9 broad stripes are superimposed on this global expression (data not shown). As zygotic expression takes over, *arrow* mRNA stripes become more accentuated (Fig. 2d), such that by stage 13 the highest levels are posterior to the En domain, and the lowest levels are just anterior to the En domain, in the region signalled by Wg at this stage (Fig. 2d, dots). Similarly, in leg discs *arrow* transcription is lowest in cells expressing Wg (data not shown). This suggests that *arrow* transcription is negatively regulated by Wg signalling, in a manner similar to the Frizzled receptors Fz and DFz2 (refs 3, 4), but unlike the signal transducers Dsh and Arm, which are both uniformly expressed^{5,6}. Although the significance of this transcriptional regulation is unclear, it is unlikely that it is relevant to signalling in the embryo, as global maternal *arrow* contribution suffices for most epidermal Wg signalling events. Polymerase chain reaction with reverse transcription (RT-PCR) analysis shows also that *arrow* mRNA is expressed in S2 cells (data not shown), consistent with these cells being able to respond to Wg once transfected with DFz2 (ref. 3).

arrow encodes a protein of 1,678 amino acids exhibiting a striking sequence conservation to the mammalian LDL-receptor related proteins LRP5 and LRP6 (Fig. 3; 71% similarity, 40% identity^{7–10}; a complete alignment is posted at <http://www.med.upenn.edu/~cellbio/DiNardo.html>; see also Supplementary Information). Arrow is a putative type I transmembrane protein containing four epidermal growth factor (EGF)-like repeats, each preceded by six YWTD spacer domains. Between the EGF/YWTD region and the membrane are LDL-receptor (LDLR) type A repeats. Both classes of repeats have been implicated in ligand binding in other contexts. For example, the LDLR repeats in the LDL receptor are central to binding the LDL particle, whereas EGF-like repeats contribute to ligand binding by some LRPs¹¹. The cytoplasmic tail contains proline-rich regions that are potential targets for SH3-domain-containing proteins, and the sequence YKII is a potential signal for internalization. Sequencing of two *arrow* alleles indicates that they are most probably null mutants. In *arr*² a stop codon is introduced at position 753, between EGF-like repeats 2 and 3 (Fig. 3, asterix;

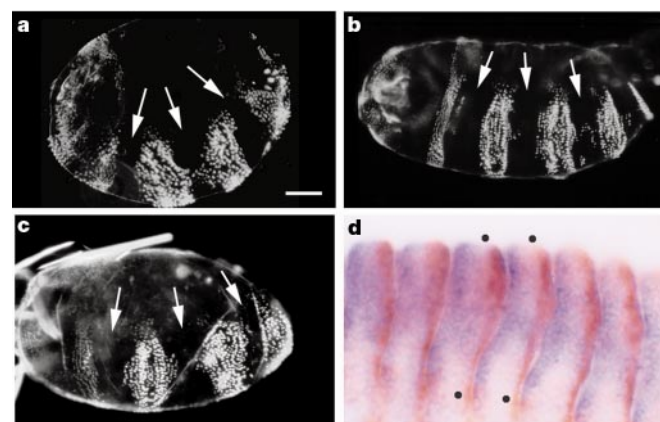


Figure 2 *arr* functions downstream of *wg* but upstream of *dsh*. Ventral (b) or ventrolateral (a, c), *arr*^{null} UAS-*Arr*; Prd-GAL4; smooth cuticle (arrows), indicating restored Wg signalling, now interrupts the lawn of denticles. b, *arr*^{null}, UAS-Dsh/Prd-GAL4; overexpression of Dsh in alternate segments produces smooth cuticle (arrows), indicating restored Wg signalling to this *arr*^{null} embryo. Similar results were obtained with activated Armadillo (not shown). c, *wg*^{ts4}, UAS-Dsh/Prd-GAL4; overexpression of Dsh in alternate segments produces smooth cuticle (arrows), showing that excess Dsh activates signalling in a ligand-independent manner. d, Arrow mRNA *in situ*. WT, stage 13; Arrow mRNA (blue) is expressed at high levels just posterior to En-expressing cells (brown), and at lower levels in regions signalled by Wg (dots). Ventral midline is along panel bottom. Scale bars, 77 μ m (a–c); 30 μ m (d).

data not shown). Another allele, *arr*¹³⁴, was generated by imprecise P-element excision and remove sequences that include the translation start and part of the signal peptide up to codon 57 (in addition to P-element internal sequences).

Arrow, LRP5 and LRP6 form a distinct group of LRPs based on overall sequence conservation (data not shown, see <http://www.med.upenn.edu/~cellbio/DiNardo.html>). The conservation among Arrow, LRP5 and LRP6 is not restricted to the repeats, but is found throughout the proteins. In addition, the blocks of EGF-like and LDL-receptor repeats are in reversed order as compared with other LRP members. Finally, the EGF-like repeats of Arrow are slightly more closely related to those of LRP5 and LRP6 than to those in other LRP proteins. The same is true for the LDL receptor repeats (data not shown). These comparisons indicate a significant functional conservation between *Drosophila* Arrow and human LRP5 and LRP6.

As Arrow appears not to be involved in production of Wg, Arrow might be required in cells receiving Wg input. To test this, we examined the expression of Wg target genes in tissues mosaic for *arrow* function. If Arrow is required to produce active signal, then *arrow* mutant cells will be rescued by signal produced from adjacent Arrow-positive cells. On the other hand, if Arrow is necessary in the receiving cell, then all the *arrow* mutant cells will exhibit a defect in Wg target gene expression. Wg is expressed along the dorso-ventral boundary of the wing disc. Secreted from these cells, Wg activates the transcription of Distalless (Dll), and acts in a concentration-dependent manner to pattern the wing^{12,13}. In the wing pouch, none of the *arrow* mutant cells within a clone expresses Dll (Fig. 4a–d). As mutant cells at the edge of the clone are not rescued by Wg secreted from their wild-type neighbours, *arrow* function is necessary in the cell receiving Wg. This is supported by analysis in the leg disk, where Wg is expressed in a ventral quadrant, and specifies ventral and ventro-lateral fates. Proper patterning also requires that *dpp*, a secreted factor necessary for dorsal and dorso-lateral fates, is repressed in cells receiving Wg. If Wg signal transduction is blocked, as in clones of cells mutant for *dsh* (refs 14–17), *dpp* expression is de-repressed and patterning is disrupted. Similarly, *dpp* expression is de-repressed in clones of cells mutant for *arrow* in the ventral anterior quadrant (Fig. 4e–h). These analyses show that *arrow* is necessary in cells responding to Wg input, for both positive and negative gene regulation.

Having established that *arrow* acts in the responding cell, we ordered the requirement for *arrow* relative to the intracellular transduction cascade. Smooth cuticle is restored to *arr*^{null} embryos in alternate segments when Dsh is expressed using Prd-GAL4, indicating rescue of Wg signal transduction (Fig. 2b). This contrasts with the overexpression of Wg, which had no effect in *arr*^{null} embryos (data not shown). These data suggest that Arrow acts downstream of Wg but upstream of Dsh, as signalling, once activated by Dsh, no longer requires Arrow. It remains possible that Arrow might normally act as a scaffold and concentrate Dsh to an appropriate subcellular location; in this model, flooding the cell with Dsh simply bypasses the requirement for Arrow. In either case, Arrow is unlikely to act in a pathway parallel to Wg signal transduction as (1) the *arrow* mutant phenotype is rescued by

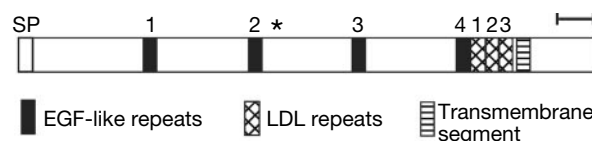


Figure 3 Arrow is a member of the LDL-receptor protein family. Arrow contains four EGF-like repeats (class B2; filled boxes), three LDL receptor repeats (type A; cross-hatching), a transmembrane domain (hatched) and a cytoplasmic tail. SP, signal peptide; bar, 100 amino acids; asterix, stop codon in *arr*².

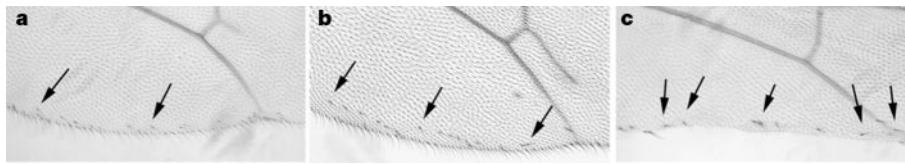


Figure 5 Ectopic expression of Arrow mimics ligand-dependent pathway activation. Adult wings, magnification of posterior margin. Arrows and arrowheads indicate some ectopic margin bristles. **a**, Overexpression of DFz2 generates ectopic margin bristles, in a Wg-dependent fashion¹⁹. **b**, Similarly, overexpression of Arrow in clones also produces ectopic bristles in the vicinity of the margin. **c**, Wing with *arrow* clones, simultaneously

overexpressing DFz2. Loss of *arrow* in clones causes notching, with loss of margin bristles; yet, overexpression of DFz2 (as in **a**), rescues neither any *arrow* mutant tissue nor mutant bristles (marked by yellow). A large notch resulting from loss of *arrow* function is apparent, yet all bristles found nearby (arrows) are generated from wild-type tissue.⁸

Dsh, a canonical Wg signal transducer, and (2) loss of *arrow* function blocks signalling even when excess Wg is presented. In addition, the fact that the *arrow* phenotype is not suppressed by excess Wg distinguishes *arrow* from the genes involved in proteoglycan-assisted presentation of the Wg ligand. Each of the mutants affecting this step, *sugarless*, *sulfateless* and *dally*, is substantially suppressed by providing excess Wg ligand, showing that glycosaminoglycans are not essential for reception of the signal but only increase the efficiency of ligand presentation to the receptor (ref. 18 and references therein). Together, these data argue that Arrow is absolutely essential for Wg to signal, and are consistent with a role for Arrow in reception rather than presentation of signal.

The mosaic and epistasis analyses place the requirement for Arrow activity upstream of Dishevelled in responding cells. Because *arrow* encodes a putative transmembrane protein, epistasis tests between Arrow and Fz proteins would be of interest, but no activating mutations exist in either Fz or Arrow that would lead to signal transduction in the absence of ligand. There are, however, several suggestive similarities between Fz proteins and Arrow. Arrow and DFz2 transduction is modulated similarly, whereas that of the Dsh and Arm signal transducers is not. In addition, ectopic DFz2 expression can mildly sensitize cells to Wg signalling, for example in the wing, where overexpression of DFz2 produces

ectopic margin bristles—a Wg-dependent cell type¹⁹ (Fig. 5a). This potentiation of Wg signalling is ligand-dependent, and ectopic bristles are only found near the wing margin, which is a source of high levels of Wg¹⁹. As is expected for a gene essential for Wg signalling, loss of *arrow* function in clones leads to loss of margin (data not shown; see Fig. 5c), similar to that observed for *fz* *Dfz2* clones²⁰. We find that overexpression of Arrow also produces ectopic bristles near the wing margin (Fig. 5b). Thus, *arrow* is required for Wg-dependent signalling at the margin, and can potentiate Wg signalling in a manner similar to that of DFz2. We also find that the potentiation of signalling caused by excess DFz2 depends on *arrow* function, as shown by inducing *arrow* mutant clones while overexpressing DFz2 (Fig. 5c). If excess DFz2 were able to bypass Arrow and restore signalling on *arrow* mutant cells then these cells should retain the ability to form wing margin and produce marked *arrow* mutant margin bristles, neither of which was observed. This result is consistent either with Arrow functioning after DFz2 engages ligand, or with Arrow and the Fz class proteins functioning as co-receptors, although this epistasis must be confirmed when activating mutations become available.

The extensive homology between Arrow and both mouse and human LRP5 and LRP6, indicates that the role we ascribe to Arrow in Wg signalling may extend to these LRPs for Wnt signalling. Indeed, an insertion in the mouse LRP6 gene has been identified that leads to several Wnt-like phenotypes²¹. Our work suggests that Arrow, and by extension LRP5 and LRP6, have specific roles for the Wg/Wnt pathway, as both Dpp and Hh can signal to *arrow* null cells (S.T.D. and S.D., unpublished data)²². We find that *arrow* null mutant cells do not survive as well as their wild-type sister cells when twin spot clones are made, suggesting some role in viability (S.T.D. and S.D., unpublished data) (see also Fig. 5c). Whether this can be attributed to a deficit in Wnt signalling is not known, although cells doubly mutant for *fz* and *dfz2* also do not survive²⁰.

Our simplest model is that Arrow and a Fz-class protein act together as a receptor complex. Some support for this is presented in an accompanying paper²³. Alternatively, Arrow might assist in recycling the Fz receptors to the plasma membrane after Wg/Wnt ligand binding and internalization, thereby providing unoccupied receptors to allow efficient, extended signalling. This idea is suggested by the fact that Arrow is related to LDL receptors, which are sometimes involved in recycling proteins from the plasma membrane. If this were the case, then overexpression of Fz proteins should suppress the need for *arrow* function by supplying excess nascent receptors. However, overexpressing DFz2 at the wing margin did not suppress the Wg signalling defect caused by loss of *arrow* function.

As several components required for Wg signalling, such as Dsh, Sgg and Fz, are also necessary for the determination of tissue polarity (reviewed in ref. 24), we tested whether Arrow also acts in this process. The only tissue where we observed polarity defects is the eye. However, the role for Arrow in polarity in this tissue is probably indirect, reflecting an early, global role for a Wnt pathway upstream of planar polarity determination²². In other tissues, Arrow is not involved in tissue polarity. For example, *arr* mutant clones in

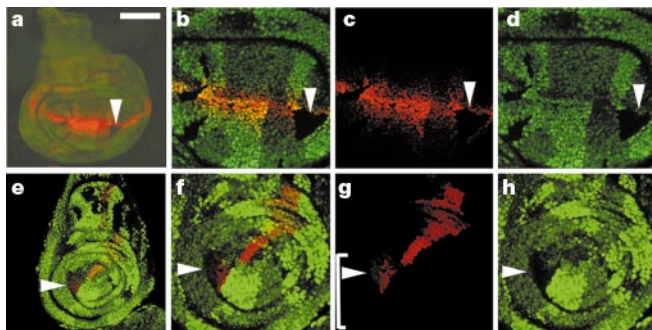


Figure 4 *arrow* is required cell-autonomously for Wg target gene expression. **a–h**, Homozygous *arr*^{null} clones marked by the absence of pi-Myc marker (green). Heterozygous tissue stains at intermediate levels, whereas the homozygous wild-type twin spot containing two copies of pi-Myc stains more brightly. **a**, Dll–LacZ expression (red) in wing disc, magnified in **b–d**; dorsal is up. Dll expression is lost in *arr*^{null} clones at the dorso–ventral boundary (arrowhead, **a–c**). Note the stronger Dll expression in wild-type twin spots compared with heterozygous tissue, indicating an *arrow* gene-dosage effect on the efficiency of Wg signal transduction. **e**, Dpp–LacZ expression (red), magnified in **f–h**; anterior is to the left, dorsal up. Dpp expression is induced in anterior compartment cells by Hh signalling from the posterior compartment. Wg signalling, active in ventral anterior sectors, represses *dpp*, so that high-level Dpp expression is normally restricted to a stripe of cells located dorsally along the antero–posterior compartment border. An *arr* clone (arrowhead) in the ventral, anterior sector (Fig. 4g, bracket), near the compartment boundary, now expresses Dpp–LacZ ectopically. Expression is weaker in mutant cells more distant from the Hh-expressing source, similar to the graded endogenous Dpp expression observed dorsally. Scale bars, 80 μ m (**a**); 35 μ m (**b–d**); 65 μ m (**e**); 35 μ m (**f–h**).

the dorsal leg exhibit no such defects (data not shown). This result also shows that the Fz receptor is available to respond to tissue polarity signals in the absence of *arrow* function and one would therefore expect it to be at the cell surface in *arrow* mutant cells. This finding suggests that Arrow does not merely act as a chaperone to guide Fz-class proteins to the cell surface. This again is consistent with our model that Fz and Arrow have a similar role in the reception of the Wg signal and that they may both be part of a single receptor complex. Recent work raises the possibility that other LRPs are part of a signal reception complex, just as we postulate for Arrow. The VLDL and ApoE receptors (both LRPs) are essential during mouse cerebellar development, where they bind the ligand, Reelin, while intracellularly binding to, and inducing the phosphorylation of, the adapter protein, Disabled-1 (reviewed in ref. 25). In addition to the LRPs, a second receptor—the Cadherin-related neuronal receptor—also binds Reelin, while its intracellular domain associates with the Fyn tyrosine kinase. This suggests that as the two receptor subunits bind Reelin, two proteins are brought into proximity inside the cell, Disabled-1 and Fyn²⁶. Perhaps Arrow and the Fz proteins similarly bind ligand and consequently bring together proteins in the cytoplasm that initiate Wg/Wnt signalling. □

Methods

Two cDNAs, corresponding to CT15575, isolated from a 0–4h library²⁷ contained an open reading frame preceded by stop codons in all frames. Two ATGs are found 5' to a putative signal peptide. For UAS–Arrow, cDNA F1 from pNB40 was cloned into pUASg as a *XhoI*/partial *NorI* fragment.

*arr*² germline clones²⁸ were induced in HS-Flp/+; FRT42B *arr*²/FRT42B *ovo*^{D1} females mated to Df(2R)8–104/CyO or Df(2R)AA1/CyO Hb-lacZ males, and null embryos were identified by absence of Hb-lacZ staining. Disc clones were induced in *y w* hs-Flp; *dpp*¹⁰⁶³⁸ (or Dll–lacZ) FRT43D *arr*²/FRT43D pi-Myc (45F, 47F) larvae²⁹. Confocal projections were captured with LSM510 software on a Zeiss axiovert microscope. For adult wings, MS1096–GAL4 was used to express UAS–DFz2 (ref. 19), or UAS-FRT-white+UAS-FRT-Arrow. Flies analysed were of genotype *y w* hs-flipase MS1096/ *y w*; UAS-DFz2/+ *y w* hs-flipase MS1096/ *w*; Df(2)8–104 UAS-FRT white+ FRT-Arrow /+, where the white+ FRT cassette was removed in clones after heat-shock-induced flipase recombination²². Marked *arr* clones were induced while overexpressing DFz2 in *y w* hs-flipase MS1096/ *y w*; FRT42D *arr*²/FRT42D *y+* *w+*; UAS-DFz2/+ flies.

Received 21 January; accepted 2 August 2000.

- Wodarz, A. & Nusse, R. Mechanisms of Wnt signalling in development. *Annu. Rev. Cell Dev. Biol.* **14**, 59–88 (1998).
- Dougan, S. T. & DiNardo, S. *wingless* generates cell type diversity among *engrailed* expressing cells. *Nature* **360**, 347–350 (1992).
- Bhanot, P. et al. A new member of the frizzled family from *Drosophila* functions as a *Wingless* receptor. *Nature* **382**, 225–230 (1996).
- Müller, H., Samanta, R. & Wieschaus, E. *Wingless* signalling in the *Drosophila* embryo: zygotic requirements and the role of the frizzled genes. *Development* **126**, 577–586 (1999).
- Klingensmith, J., Nusse, R. & Perrimon, N. The *Drosophila* segment polarity gene *dishevelled* encodes a novel protein required for response to the *wingless* signal. *Genes Dev.* **8**, 118–130 (1994).
- Riggelman, B., Schedl, P. & Wieschaus, E. Spatial expression of the *Drosophila* segment polarity gene *armadillo* is posttranscriptionally regulated by *wingless*. *Cell* **63**, 549–560 (1990).
- Hey, P. J. et al. Cloning of a novel member of the low-density lipoprotein receptor family. *Gene* **216**, 103–111 (1998).
- Kim, D. H. et al. A new low density lipoprotein receptor related protein, LRP5, is expressed in hepatocytes and adrenal cortex, and recognizes apolipoprotein E. *J. Biochem. (Tokyo)* **124**, 1072–1076 (1998).
- Dong, Y. et al. Molecular cloning and characterization of LR3, a novel LDL receptor family protein with mitogenic activity. *Biochem. Biophys. Res. Commun.* **251**, 784–790 (1998).
- Chen, D., Lathrop, W. & Dong, Y. Molecular cloning of mouse Lrp7(Lr3) cDNA and chromosomal mapping of orthologous genes in mouse and human. *Genomics* **55**, 314–321 (1999).
- Brown, M. S., Herz, J. & Goldstein, J. L. LDL-receptor structure. Calcium cages, acid baths and recycling receptors. *Nature* **388**, 629–630 (1997).
- Zecca, M., Basler, K. & Struhl, G. Direct and long-range action of a *wingless* morphogen gradient. *Cell* **87**, 833–844 (1996).
- Neumann, C. J. & Cohen, S. M. Long-range action of *Wingless* organizes the dorsal–ventral axis of the *Drosophila* wing. *Development* **124**, 871–880 (1997).
- Brook, W. J. & Cohen, S. M. Antagonistic interactions between *wingless* and *decapentaplegic* responsible for dorsal–ventral pattern in the *Drosophila* leg. *Science* **273**, 1373–1377 (1996).
- Jiang, J. & Struhl, G. Complementary and mutually exclusive activities of *decapentaplegic* and *wingless* organize axial patterning during *Drosophila* leg development. *Cell* **86**, 401–409 (1996).
- Penton, A. & Hoffmann, F. M. *Decapentaplegic* restricts the domain of *wingless* during *Drosophila* limb patterning. *Nature* **382**, 162–164 (1996).
- Heslop, T. R., Theisen, H., Walker, H. & Marsh, J. L. Shaggy and *dishevelled* exert opposite effects on *Wingless* and *Decapentaplegic* expression and on positional identity in imaginal discs. *Development* **124**, 1069–1078 (1997).
- Perrimon, N. & Bernfield, M. Specificities of heparan sulphate proteoglycans in developmental processes. *Nature* **404**, 725–728 (2000).

- Cadigan, K. M., Fish, M. P., Rulifson, E. J. & Nusse, R. *Wingless* repression of *Drosophila* frizzled 2 expression shapes the *Wingless* morphogen gradient in the wing. *Cell* **93**, 767–777 (1998).
- Chen, C. & Struhl, G. *Wingless* transduction by the frizzled and frizzled2 proteins of *Drosophila*. *Development* **126**, 5441–5452 (1999).
- Pinson, K., Brennan, J., Monkley, S., Avery, B. & Skarnes, W. C. An LDL-receptor-related protein, regulates Wnt signalling in mice. *Nature* **407**, 535–538 (2000).
- Wehrli, M. & Tomlinson, A. Independent regulation of anterior/posterior and equatorial/polar polarity in the *Drosophila* eye; evidence for the involvement of Wnt signalling in the equatorial/polar axis. *Development* **125**, 1421–1432 (1998).
- Tamai, K. et al. LDL-receptor-related proteins in Wnt signal transduction. *Nature* **407**, 530–535 (2000).
- Axelrod, J. D., Miller, J. R., Shulman, J. M., Moon, R. T. & Perrimon, N. Differential recruitment of Dishevelled provides signalling specificity in the planar cell polarity and *Wingless* signalling pathways. *Genes Dev.* **12**, 2610–2622 (1998).
- Rice, D. S. & Curran, T. Mutant mice with scrambled brains: understanding the signalling pathways that control cell positioning in the CNS. *Genes Dev.* **13**, 2758–2873 (1999).
- Senzaki, K., Ogawa, M. & Yagi, T. Proteins of the CNR family are multiple receptors for Reelin. *Cell* **99**, 635–647 (1999).
- Brown, N. H. & Kafatos, F. C. Functional cDNA libraries from *Drosophila* embryos. *J. Mol. Biol.* **203**, 425–437 (1988).
- Chou, T. B., Noll, E. & Perrimon, N. Autosomal P[ovoD1] dominant female-sterile insertions in *Drosophila* and their use in generating germ-line chimeras. *Development* **119**, 1359–1369 (1993).
- Wehrli, M. & Tomlinson, A. Epithelial planar polarity in the developing *Drosophila* eye. *Development* **121**, 2451–2459 (1995).
- Nüsslein-Volhard, C., Wieschaus, E. & Kluding, H. Mutations affecting the pattern of the larval cuticle in *Drosophila melanogaster* I Zygotic loci on the second chromosome. *Roux's Arch. Dev. Biol.* **193**, 267–282 (1984).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or from <http://www.med.upenn.edu/~cellbio/DiNardo.html> or as paper copy from the London editorial office of *Nature*.

Acknowledgements

G. Campbell, G. Struhl, F. Diaz-Benjumea, S. Cohen, R. Goto-Mandeville, E. Bieschke and B. Calvi provided insightful suggestions and help along the way. The free exchange of information with the Skarnes lab is acknowledged. The manuscript was improved by comments from N. Erdeniz, P. Klein, B. Wilder and the DiNardo lab. Material provided by J. Szidonya, the Bloomington Stock Center and Berkeley *Drosophila* Genome Project was of great importance. Supported by the Swiss National Science Foundation (M.W.), NIH (A.T. and S.D.) and American Cancer Society (S.D.).

Correspondence and requests for materials should be addressed to S.D. (e-mail: sdinardo@mail.med.upenn.edu).

LDL-receptor-related proteins in Wnt signal transduction

Keiko Tamai*, Mikhail Semenov*, Yoichi Kato*, Rebecca Spokony†, Chunming Liu*, Yu Katsuyama*, Fred Hess‡, Jean-Pierre Saint-Jeannet† & Xi He*

* Division of Neuroscience, Children's Hospital, Department of Neurology, Harvard Medical School, 300 Longwood Avenue, Boston, Massachusetts 02115, USA

† Department of Animal Biology, School of Veterinary Medicine, University of Pennsylvania, 3800 Spruce Street, Philadelphia, Pennsylvania 19104, USA

‡ Department of Human Genetics, Merck Research Laboratories, PO Box 4, West Point, Pennsylvania 19486, USA

The Wnt family of secreted signalling molecules are essential in embryo development and tumour formation¹. The Frizzled (Fz) family of serpentine receptors function as Wnt receptors^{2–10}, but how Fz proteins transduce signalling is not understood. In *Drosophila*, *arrow* phenocopies the *wingless* (DWnt-1) phenotype¹¹, and encodes a transmembrane protein¹¹ that is homologous to two members of the mammalian low-density lipoprotein receptor (LDLR)-related protein (LRP) family, LRP5 and LRP6 (refs 12–15). Here we report that LRP6 functions as a co-receptor for Wnt signal transduction. In *Xenopus* embryos, LRP6 activated

Wnt-Fz signalling, and induced Wnt responsive genes, dorsal axis duplication and neural crest formation. An LRP6 mutant lacking the carboxyl intracellular domain blocked signalling by Wnt or Wnt-Fz, but not by Dishevelled or β -catenin, and inhibited neural crest development. The extracellular domain of LRP6 bound Wnt-1 and associated with Fz in a Wnt-dependent manner. Our results indicate that LRP6 may be a component of the Wnt receptor complex.

Human LRP5 and LRP6 share 71% amino-acid identity^{12–15}, and together with Arrow, form a distinct subgroup of the LRP family¹¹. Arrow, LRP5 and LRP6 each contain an extracellular domain with EGF (epidermal growth factor) repeats and LDLR repeats, followed by a transmembrane region and a cytoplasmic domain lacking recognizable catalytic motifs^{11–15}. An *lrp6* mutation in mice results in pleiotropic defects recapitulating some, but not all, of the *wnt* mutant phenotype¹⁶. To study LRP5/LRP6 involvement in Wnt signalling, we have examined their function in Wnt-induced axis and neural crest formation in *Xenopus* embryos.

Wnt/ β -catenin signalling induces dorsal axis formation through activation of responsive genes, including nodal-related 3 (*Xnr3*) and siamois (*sia*) (reviewed in ref. 17). Ventral injection of LRP6 RNA into four-cell stage embryos resulted in dorsal axis duplication in a dose-dependent manner (Fig. 1a, b). In animal pole explants, LRP6 induced *Xnr3*/*sia*, but not brachyury (*Xbra*) expression (Fig. 1d), which is activated by mesoderm inducers like activin or basic fibroblast growth factor (bFGF)¹⁷. These results indicate that over-expression of LRP6 may specifically activate Wnt signalling. To examine whether LRP6 mediates Wnt effect, we co-injected RNAs

for LRP6 and Wnt-5a. Neither Wnt-5a nor a low dose of LRP6 alone exhibited any effect, but Wnt-5a plus LRP6 synergistically induced axis duplication and ectopic *Xnr3* expression in the embryo (Fig. 1a–c), and activated *Xnr3*/*sia* in explants (Fig. 1d). Synergy was also observed between Wnt-5a and hFz5 as reported⁴, and between LRP6 and hFz5 (Fig. 1e). Although LRP5 alone did not induce axes, co-injecting LRP5 and Wnt-5a did (Fig. 1a). LDLR alone or in combination with Wnt-5a did not induce axes or *Xnr3*/*sia* (Fig. 1a and d). Although Wnt-5a-hFz5 can induce complete axes including head and the notochord⁴, Wnt-5a-LRP6 or LRP6 alone (higher doses) induced trunk axis with muscle and neural tissues but lacking head and the notochord (Fig. 1b). This may be explained by quantitative or qualitative differences between Wnt-5a/LRP6 and Wnt-5a/hFz5 co-injections.

Because ectopic Wnt expression enhances, whereas lack of Wnt signalling inhibits, neural crest formation^{18–22}, we further analysed the effect of LRP6 on neural crest development. LDLR injection had no effect, but LRP6 expression significantly expanded neural crest progenitors in the injected half of the embryo, as determined by the expression of a crest-specific marker, slug (Fig. 2a, b). Thus, over-expression of LRP6 also mimicked Wnt signalling during neural crest formation.

To distinguish whether LRP6 functions in Wnt-responding or Wnt-producing cells, we injected Wnt-5a and LRP6 separately into neighbouring blastomeres at the four-cell stage (Fig. 3a, insert). Induction of secondary axes in embryos and of *Xnr3*/*sia* in explants occurred even when Wnt-5a and LRP6 were expressed in different cells (Fig. 3a, b). Therefore, LRP6 is probably involved in respond-

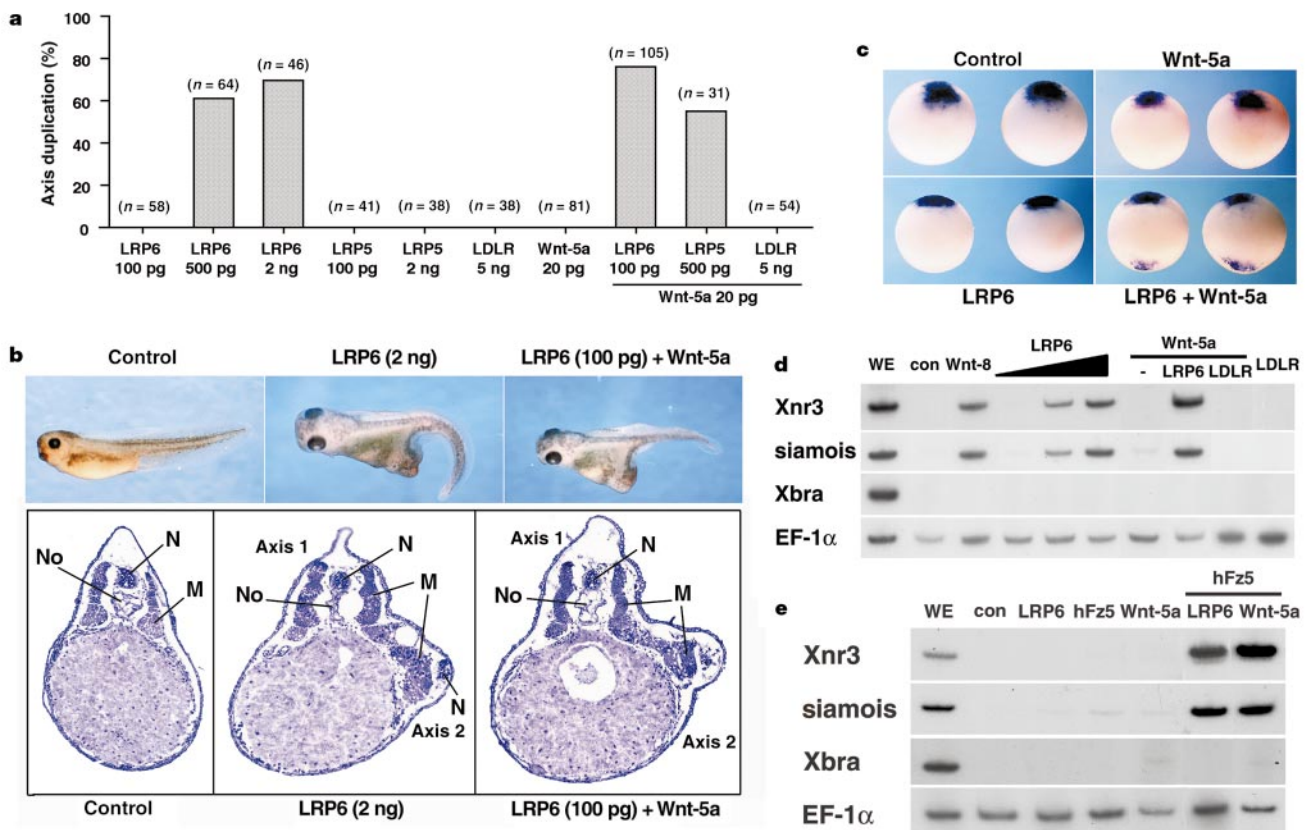


Figure 1 LRP6/LRP5 in Wnt signalling. **a**, Ventral injection of LRP6 RNA (0.5 or 2 ng), or co-injection of RNAs for *Xenopus* Wnt-5a (20 pg) plus LRP6 (100 pg) or LRP5 (500 pg) induced axis duplication. LDLR RNA (5 ng or 1 ng; not shown) alone or co-injected with Wnt-5a did not induce axes. *n*, numbers of embryos scored. **b**, LRP6 (2 ng) or Wnt-5a (20 pg) plus LRP6 (100 pg) induced trunk duplication lacking head and the notochord. Top, stage-40 embryos; bottom, histology on cross-sections. M, muscle; N, neural

tissues; No, notochord. **c**, Ventral co-injection of Wnt-5a (20 pg) and LRP6 (100 pg) activated ectopic *Xnr3* expression (stage 10.5), assayed by *in situ* hybridization (*n* = 10–12). **d, e**, Synergistic induction of *Xnr3*/*sia* in animal pole explants (stage 10.5) by LRP6 (100 pg) plus Wnt-5a (20 pg) or hFz5 (400 pg), assayed by RT-PCR. The RNA amount injected is as in **a** except for *Xenopus* Wnt-8 (10 pg). EF-1 α , loading control. WE, whole embryos (stage 10.5); con, explants from uninjected embryos.

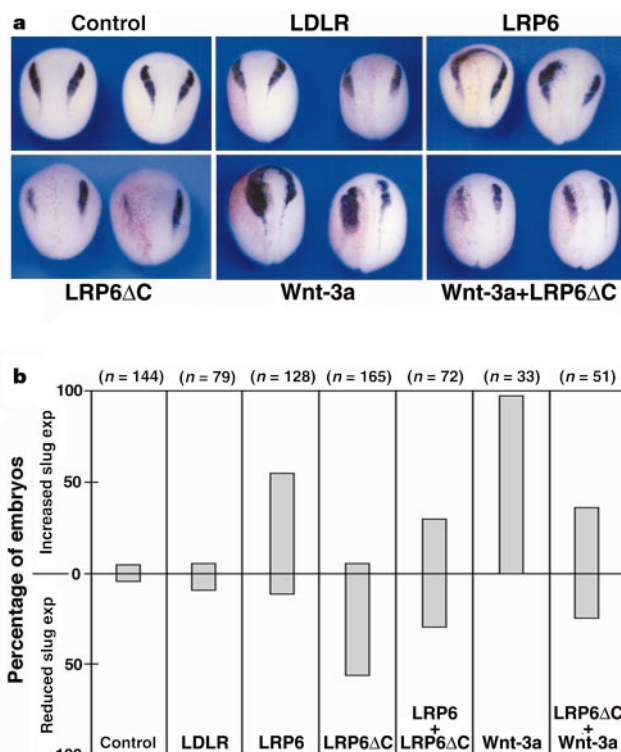


Figure 2 LRP6 in neural crest formation. **a**, Slug expression (stage 15–20) examined by *in situ* hybridization. LRP6 RNA (2 ng) or *Xenopus* Wnt-3a DNA (100 pg) increased, whereas LRP6 Δ C (2 ng) inhibited slug expression in the injected half (the left side,

labelled by red-coloured staining derived from co-injected β -galactosidase RNA). LRP6 Δ C antagonized Wnt-3a, whose DNA was used to mimic zygotic Wnt expression. LDLR (2 ng) had no effect. **b**, Summary of the *in situ* hybridization results.

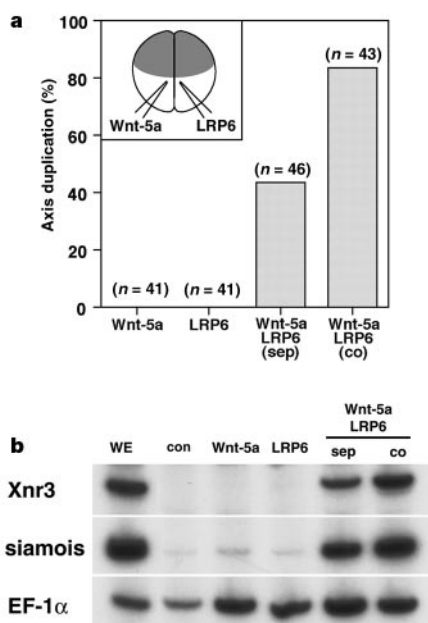


Figure 3 LRP6 functions in Wnt-responding cells. RNAs for Wnt-5a (100 pg) and LRP6 (100 pg) were separately injected (**a**, insert) into two neighbouring blastomeres in the ventral equatorial region (**a**) or in the animal pole (**b**). sep, separate injection; co, Wnt-5a and LRP6 co-injected into a single blastomere for comparison. **a**, Percentage of embryos with axis duplication. **b**, Xnr3/sia induction in animal pole explants. Separate injections were less effective than co-injections (**a**, **b**), probably because of limited diffusion of Wnt-5a protein. EF-1 α , loading control. WE, whole embryos (stage 10.5); con, explants from uninjected embryos.

ing to, rather than in enhancing, the production or secretion of the Wnt ligand.

To inhibit the function of the endogenous LRP6 which is expressed maternally and throughout embryogenesis (Fig. 4f), we generated LRP6 Δ C which has most of its cytoplasmic domain deleted. LRP6 Δ C did not, either alone or in combination with Wnt-5a, induce axes or activate Xnr3/sia (Fig. 4a, b), but inhibited axis duplication and Xnr3/sia induction by the wild-type LRP6 (Fig. 4a, b). This inhibition was counteracted by an increasing amount of co-injected LRP6 (Fig. 4a). These data suggest that LRP6 Δ C is a dominant interfering mutant for LRP6 or related molecules, and that LRP6 cytoplasmic domain is required for Wnt signalling. LRP6 Δ C inhibited Xnr3/sia induction by several Wnt molecules, including Wnt-1, Wnt-2, Wnt-3a and Wnt-8 (Fig. 4c). LRP6 Δ C also inhibited Wnt-5a signalling through hFz5 (Fig. 4c), showing that hFz5, and probably other endogenous Fz molecules mediating Wnt-1 or Wnt-8 signalling, depend on LRP6 or related proteins. LRP6 Δ C did not affect Xbra induction by activin or bFGF (Fig. 4d), and thus interferes specifically with Wnt signalling.

LRP6 Δ C injected dorsally at the four-cell stage did not perturb the endogenous axis formation (data not shown). This may mean that the dorsal β -catenin pathway is activated by mechanisms other than Wnt stimulation¹⁷; alternatively, the dorsal Wnt–Fz signalling may occur early²³ before LRP6 Δ C can interfere. However, LRP6 Δ C inhibited neural crest development as examined by slug expression, and suppressed ectopic crest formation induced by Wnt-3a DNA (Fig. 2a and b). Furthermore, co-injection of LRP6 rescued LRP6 Δ C inhibition of crest formation (Fig. 2b). Thus, LRP6 or a related molecule is required for Wnt-dependent neural crest formation *in vivo*.

In the current model of Wnt/ β -catenin signalling, Wnt stimulation of a Fz receptor activates the intracellular protein Dishevelled (Dsh or Dvl), thereby antagonizing the inhibitory action of the Axin/GSK-3 complex and stabilizing β -catenin, which together

with the transcription factor TCF/LEF activates responsive genes¹. To position LRP6 in this cascade, we tested relationships between LRP6 and other Wnt signalling components. LRP6 Δ C inhibited Xnr3/sia induction by Wnts, but not by Dsh or β -catenin (Fig. 4c), suggesting that LRP6 Δ C interfered with Wnt signalling upstream of Dsh function. Supporting this epistasis, LRP6 induction of Xnr3/sia was antagonized by Axin²⁴, by a dominant-negative TCF, Δ NTCF²⁵, and by a dominant-negative Dsh, mDvl2-DIX (Fig. 4e). LRP6 activity was also inhibited by frizzled-related protein (FRP), a secreted Wnt antagonist²⁶ (Fig. 4e), implying that LRP6 activation of Wnt signalling relied on endogenous Wnt molecules. Alternatively, FRP may directly inhibit LRP6. Thus, LRP6 acts between the extracellular Wnt, FRP and intracellular Dsh, possibly as a co-receptor for Wnt molecules.

To function as a Wnt co-receptor, LRP6 should bind Wnt or Fz or both. To examine the issue, we used a secreted form of mFz8, mFz8CRD-IgG (ref. 8), which comprised the cysteine-rich domain (CRD) of mFz8 N-terminal extracellular region fused with the immunoglobulin- γ (IgG) Fc epitope, and a secreted LRP6N-Myc, which consisted of the LRP6 extracellular domain tagged by the Myc epitope. We incubated mFz8CRD-IgG and LRP6N-Myc proteins with or without Wnt-1. mFz8CRD-IgG co-precipitated LRP6N-Myc only in the presence of Wnt-1; the secreted IgG fusion partner failed to do so regardless of Wnt-1 (Fig. 5A). We also performed a reciprocal precipitation using secreted LRP6N-IgG

and mFz8CRD-Myc, which we generated by swapping the two epitopes. LRP6N-IgG co-precipitated mFz8CRD-Myc, again only in the presence of Wnt-1, whereas the control IgG did not (Fig. 5B). LRP6N-IgG also co-precipitated Wnt-1-Myc (Fig. 5C), a tagged Wnt-1 protein. These results suggest that the extracellular domain of LRP6 can bind Wnt-1 and form a complex with Fz in a Wnt-dependent fashion.

We have shown that in two developmental processes dependent on the Wnt pathway in *Xenopus*—secondary axis and neural crest formation—LRP6 activates but a dominant-negative LRP6 inhibits Wnt signalling, providing compelling evidence that LRP6 is critical in Wnt signal transduction. LRP6 functions upstream of Dsh in Wnt-responding cells, synergizes with either Wnt or Fz, and importantly, is able to bind Wnt-1 and to associate with Fz in a Wnt-dependent manner. The simplest interpretation of these findings is that LRP6 is a component of the Wnt-Fz receptor complex. Genetic studies of *arrow* in *Drosophila*¹¹ and *lrp6* in mice¹⁶ strongly support this hypothesis. Our binding data also raise the possibility that Wnt-induced formation of the Fz-LRP6 complex assembles LRP6, Fz and their associated proteins, thereby initiating cytoplasmic signalling. Consistent with this notion, Wnt signal transduction requires intracellular regions of both Fz (data not shown) and LRP6, which harbours candidate protein-protein interaction motifs^{11–15}. Notably, *arrow* does not exhibit *fz* planar polarity phenotype¹¹, implying that Arrow-LRP6 may specify Wnt-Fz signalling towards

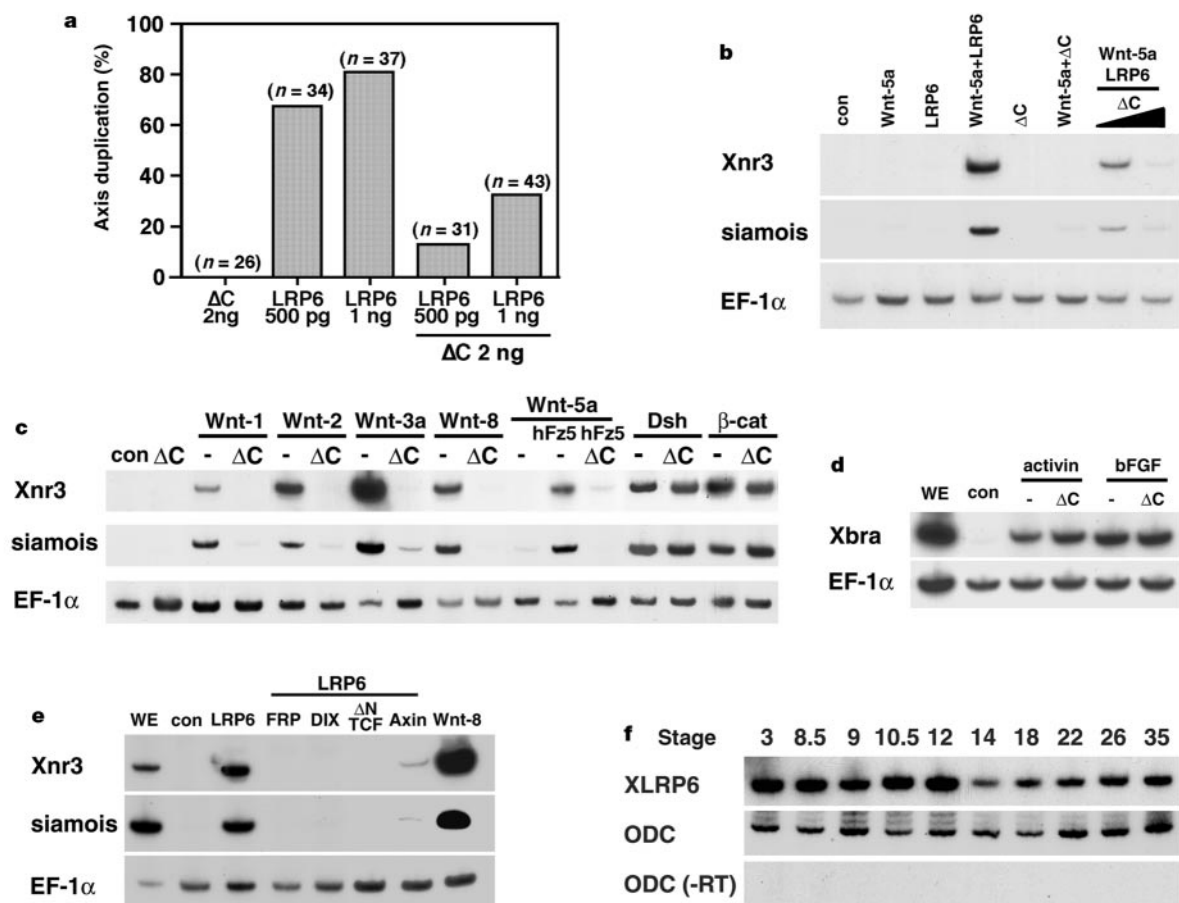


Figure 4 LRP6 functions upstream of Dsh protein and is antagonized by FRP. **a**, LRP6 Δ C (labelled Δ C, 2 ng) did not induce axes but inhibited axis induction by LRP6 (500 pg) when co-injected. An increase of LRP6 RNA (1 ng) counteracted LRP6 Δ C inhibition. **b–e**, Animal pole explant assays. **b**, LRP6 Δ C alone (2 ng) or Wnt-5a (20 pg) plus LRP6 Δ C (100 pg) did not induce Xnr3/sia. LRP6 Δ C (500 pg and 2 ng) inhibited Xnr3/sia induced by Wnt-5a (20 pg) plus LRP6 (100 pg) in a dose dependent manner. **c**, LRP6 Δ C (2 ng) inhibited Xnr3/sia induction by Wnt-1 (10 pg), Wnt-2 (40 pg), Wnt-3a (5 pg), Wnt-8 (10 pg) and by Wnt-5a (20 pg) plus hFz5 (100 pg), but not by Dsh (1 ng) or β -catenin

(100 pg). *Xenopus* Wnts were used except for mouse Wnt-1. **d**, LRP6 Δ C (2 ng) did not perturb Xbra induction by activin or bFGF (50 ng ml⁻¹). **e**, LRP6 (500 pg) induction of Xnr3/sia was inhibited by co-injected RNAs (2 ng) for FRP, a dominant-negative Dsh (DIX), Δ NTCF and Axin. **f**, *Xenopus* LRP6 is expressed maternally and throughout embryogenesis, as assayed by RT-PCR. Ornithine decarboxylase was used as a control for relative RNA amount. –RT, PCR without RT. A partial cDNA of XLRP6 showed 80% amino-acid sequence identity with the corresponding region of human LRP6. EF-1 α , loading control. WE, whole embryos (stage 10.5); con, explants from uninjected embryos.

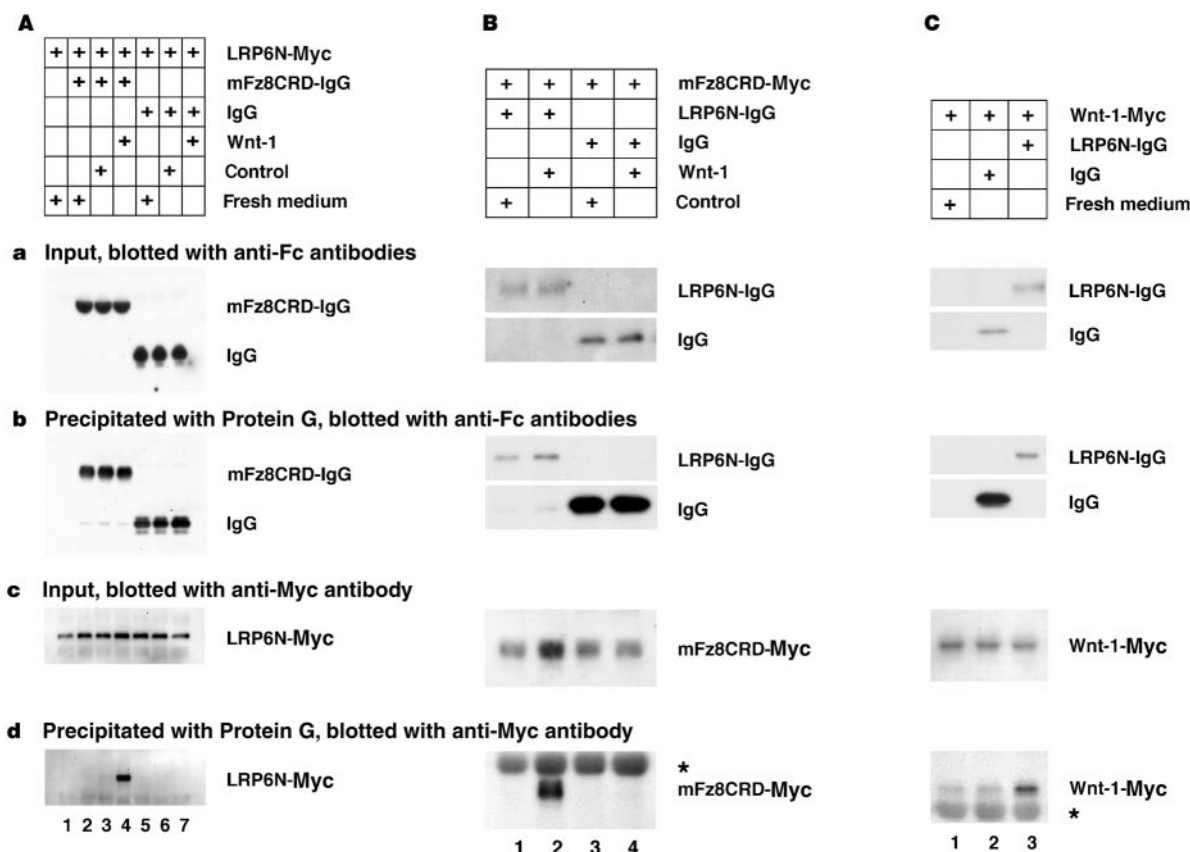


Figure 5 LRP6 extracellular domain binds Wnt-1 and complexes with mFz8CRD in the presence of Wnt-1. All components were provided as conditioned medium. **A**, mFz8CRD-IgG co-precipitated LRP6N-Myc in the presence of Wnt-1. Conditioned medium mixtures containing mFz8CRD-IgG or the control IgG (**a**) and LRP6N-Myc (**c**) were incubated in the presence of Wnt-1 conditioned medium, control conditioned medium, or fresh

medium, and precipitated with Protein G beads. Precipitates were detected for mFz8CRD-IgG or the control IgG (**b**) or LRP6N-Myc (**d**). **B**, LRP6N-IgG co-precipitated mFz8CRD-Myc in the presence of Wnt-1. Note that the epitope tags were reciprocated. **C**, LRP6N-IgG co-precipitated Wnt-1-Myc. Precipitation of LRP6N-IgG was less effective than that of IgG in (**B**, **C**). Asterisk indicates IgG from bovine serum.

the β -catenin pathway. How Fz, LRP6 and proteoglycan molecules such as Dally²⁷ interact to mediate Wnt recognition/specificity and signal transduction remains to be studied. In addition, whether other LRPs and LRP-binding proteins participate in or modulate different Wnt-Fz signalling pathways needs evaluation. The LRP family have diverse signalling functions, as shown here^{11,16} and in mammalian brain development²⁸. □

Methods

cDNA constructs

We subcloned cDNAs for human LRP5 (ref. 13), LRP6 (ref. 12) and LDLR into pCS2+. LRP6 Δ C was generated by deleting residue 1,440 onwards, 3' of an internal XhoI site. mFz8CRD-IgG and the control human IgG (Fc) have been described⁸. LRP6N-IgG and LRP6N-Myc were generated by fusing the LRP6 extracellular domain (two residues before the putative transmembrane region) with the IgG⁸ or the 6-Myc epitope (from pCS2+MT). mFz8CRD-Myc and Wnt-1-Myc comprised residues 1–178 of mFz8, and mouse Wnt-1 fused with the 6-Myc epitope, respectively. We generated mDvl2-DIX by subcloning the NcoI and XmnI fragment (encoding the DIX domain) of murine Dvl2 cDNA into pCS2+MT. mDvl2-DIX behaves as a dominant-negative reagent that blocks Wnt signalling. Details of the plasmids are available upon request.

Embryo manipulation, histology, *in situ* hybridization and explants

These procedures, and polymerase chain reaction with reverse transcription (RT-PCR), were done as described^{19,29}.

Conditioned medium, co-precipitation and immunoblotting

We produced LRP6N-IgG (apparent relative molecular mass 200K), mFz8CRD-IgG (60K), the control IgG (35K), LRP6N-Myc (200K) and mFz8CRD-Myc (45K) by transient transfection of 293T cells as described⁸. Wnt-1 or Wnt-1-Myc (70K) conditioned medium was collected from Rat-2 cells infected with viral vectors (ref. 30, and data not shown). Wnt-1 or Wnt-1-Myc biological activity was controlled using a β -catenin

stabilization assay (data not shown). LRP6N-IgG (1 ml) or mFz8CRD-IgG (0.5 ml), or a corresponding amount of the control IgG was mixed with 0.5 ml of mFz8CRD-Myc or LRP6N-Myc in the presence of 2 ml of Wnt-1, control, or fresh medium. Precipitation and immunoblotting were done similarly to ref. 8.

Received 14 June; accepted 28 July 2000.

- Wodarz, A. & Nusse, R. Mechanisms of Wnt signalling in development. *Annu. Rev. Cell Dev. Biol.* **14**, 59–88 (1998).
- Bhanot, P. *et al.* A new member of the frizzled family from *Drosophila* functions as a Wingless receptor. *Nature* **382**, 225–230 (1996).
- Yang-Snyder, J., Miller, J. R., Brown, J. D., Lai, C. J. & Moon, R. T. A frizzled homolog functions in a vertebrate Wnt signalling pathway. *Curr. Biol.* **6**, 1302–1306 (1996).
- He, X. *et al.* A member of the Frizzled protein family mediating axis induction by Wnt-5A. *Science* **275**, 1652–1654 (1997).
- Bhat, K. M. frizzled and frizzled 2 play a partially redundant role in wingless signalling and have similar requirements to wingless in neurogenesis. *Cell* **95**, 1027–1036 (1998).
- Kennerdell, J. R. & Carthew, R. W. Use of dsRNA-mediated genetic interference to demonstrate that frizzled and frizzled 2 act in the wingless pathway. *Cell* **95**, 1017–1026 (1998).
- Muller, H., Samanta, R. & Wieschaus, E. Wingless signalling in the *Drosophila* embryo: zygotic requirements and the role of the frizzled genes. *Development* **126**, 577–586 (1999).
- Hsieh, J. C., Rattner, A., Smallwood, P. M. & Nathans, J. Biochemical characterization of Wnt-frizzled interactions using a soluble, biologically active vertebrate Wnt protein. *Proc. Natl Acad. Sci. USA* **96**, 3546–3551 (1999).
- Bhanot, P. *et al.* Frizzled and Dfrizzled-2 function as redundant receptors for Wingless during *Drosophila* embryonic development. *Development* **126**, 4175–4186 (1999).
- Chen, C. M. & Struhl, G. Wingless transduction by the Frizzled and Frizzled2 proteins of *Drosophila*. *Development* **126**, 5441–5452 (1999).
- Wehrli, M. *et al.* *arrow* encodes an LDL-receptor-related protein essential for Wingless signalling. *Nature* **407**, 527–530 (2000).
- Brown, S. D. *et al.* Isolation and characterization of LRP6, a novel member of the low density lipoprotein receptor gene family. *Biochem. Biophys. Res. Commun.* **248**, 879–888 (1998).
- Hey, P. J. *et al.* Cloning of a novel member of the low-density lipoprotein receptor family. *Gene* **216**, 103–111 (1998).
- Kim, D. H. *et al.* A new low density lipoprotein receptor related protein, LRP5, is expressed in hepatocytes and adrenal cortex, and recognizes apolipoprotein E. *J. Biochem. (Tokyo)* **124**, 1072–1076 (1998).

15. Dong, Y. *et al.* Molecular cloning and characterization of LR3, a novel LDL receptor family protein with mitogenic activity. *Biochem. Biophys. Res. Commun.* **251**, 784–790 (1998).
16. Pinson, K. I., Brennan, J., Monkley, S., Avery, B. & Skarnes, W. C. An LDL-receptor-related protein mediates Wnt signalling in mice. *Nature* **407**, 535–538 (2000).
17. Harland, R. M. & Gerhart, J. Formation and function of Spemann's organizer. *Annu. Rev. Cell Dev. Biol.* **13**, 611–667 (1997).
18. Ikeya, M., Lee, S. M. K., Johnson, J. E., McMahon, A. P. & Takada, S. Wnt signalling required for expansion of neural crest and CNS progenitors. *Nature* **389**, 966–970 (1997).
19. Saint-Jeannet, J.-P., He, X., Varmus, H. E. & Dawid, I. B. Regulation of dorsal fate in the neuraxis by Wnt-1 and Wnt-3a. *Proc. Natl Acad. Sci. USA* **94**, 13713–13718 (1997).
20. Chang, C. & Hemmati-Brivanlou, A. Neural crest induction by Xwnt7B in *Xenopus*. *Dev. Biol.* **194**, 129–134 (1998).
21. LaBonne, C. & Bronner-Fraser, M. Neural crest induction in *Xenopus*: evidence for a two signal model. *Development* **125**, 2403–2414 (1998).
22. Dorsky, R. I., Moon, R. T. & Raible, D. W. Control of neural crest cell fate by the Wnt signalling pathway. *Nature* **396**, 370–373 (1998).
23. Sumanas, S., Strege, P., Heasman, J., & Ekker, S. C. The putative Wnt receptor *Xenopus frizzled-7* functions upstream of β -catenin in vertebrate dorsoventral mesoderm patterning. *Development* **127**, 1981–1990 (2000).
24. Ikegaya, L. *et al.* The mouse fused locus encodes Axin, an inhibitor of the Wnt signalling pathway that regulates embryonic axis formation. *Cell* **90**, 181–192 (1997).
25. Molenaar, M. *et al.* XTcf-3 transcription factor mediates β -catenin-induced axis formation in *Xenopus* embryos. *Cell* **86**, 391–399 (1996).
26. Finch, P. W. *et al.* Purification and molecular cloning of a secreted, Frizzled-related antagonist of Wnt action. *Proc. Natl Acad. Sci. USA* **94**, 6770–6775 (1997).
27. Perrimon, N. & Bernfield, M. Specificities of heparan sulphate proteoglycans in developmental processes. *Nature* **404**, 725–728 (2000).
28. Trommsdorff, M. *et al.* Reeler/Disabled-like disruption of neuronal migration in knockout mice lacking the VLDL receptor and ApoE receptor 2. *Cell* **97**, 689–701 (1999).
29. Kato, Y., Shi, Y. & He, X. Neuralization of the *Xenopus* embryo by inhibition of p300/CBP function. *J. Neuroscience* **19**, 9346–9373 (1999).
30. Shimizu, H. *et al.* Transformation by Wnt family proteins correlates with regulation of β -catenin. *Cell Growth Differ.* **8**, 1349–1358 (1997).

Acknowledgements

We thank M. Semenova for technical assistance; J. Heitz, J. Kitajewski, J. Nathans, S. Sokol, D. Sussman and A. Parlow (NHPP) for reagents; S. DiNardo and B. Skarnes for communication; and R. Habas, Z. He and Q. Ma for comments. X.H. acknowledges supports from Johnson and Johnson, the US Army, Susan G. Komen Foundation and the NIH. J.-P.S.-J. acknowledges supports from Johnson and Johnson and Whitehall Foundation. X.H. is a Pew Scholar and Klingenstein Fellow.

Correspondence and requests for materials should be addressed to X.H. (e-mail: he_x@hub.tch.harvard.edu). A partial *Xenopus* LRP6 cDNA sequence is deposited in GenBank under accession code AF276084.

An LDL-receptor-related protein mediates Wnt signalling in mice

Kathleen I. Pinson, Jane Brennan*, Susan Monkley*, Brian J. Avery & William C. Skarnes

Department of Molecular and Cell Biology, University of California, Berkeley, California 94720, USA

Wnt genes comprise a large family of secreted polypeptides that are expressed in spatially and tissue-restricted patterns during vertebrate embryonic development¹. Mutational analysis in mice has shown the importance of Wnts in controlling diverse developmental processes such as patterning of the body axis, central nervous system and limbs, and the regulation of inductive events during organogenesis². Although many components of the Wnt signalling pathway have been identified, little is known about how Wnts and their cognate Frizzled receptors signal to downstream effector molecules. Here we present evidence that a new member of the low-density lipoprotein (LDL)-receptor-related protein

family, LRP6 (ref. 3), is critical for Wnt signalling in mice. Embryos homozygous for an insertion mutation in the LRP6 gene exhibit developmental defects that are a striking composite of those caused by mutations in individual Wnt genes. Furthermore, we show a genetic enhancement of a Wnt mutant phenotype in mice lacking one functional copy of LRP6. Together, our results support a broad role for LRP6 in the transduction of several Wnt signals in mammals.

In a screen for recessive lethal phenotypes in mice caused by gene trap insertions in cell-surface proteins⁴, we recovered an insertion mutation that joined the first 321 amino acids of the LRP6 protein in-frame with the β geo reporter gene. Embryos homozygous for the insertion in LRP6 died at birth, and exhibited a variety of severe developmental abnormalities including a truncation of the axial skeleton, limb defects, microphthalmia and malformation of the urogenital system (Fig. 1a–c). Northern blot analysis showed a complete absence of LRP6 transcripts in LRP6^{−/−} embryos and embryonic fibroblasts (Fig. 1d). Southern blot analysis indicated a simple insertion event in which LRP6 exons downstream of the vector are retained (data not shown). On the basis of β geo reporter activity, which accurately reports endogenous gene expression⁴, the LRP6 gene appears to be expressed in all cells of the developing embryo (data not shown).

The specific developmental defects that we observed in homozygous embryos were remarkably similar to mice carrying mutations in Wnt genes, specifically Wnt-3a, Wnt-1 and Wnt-7a. Both the targeted null mutation and a classical hypomorphic allele of Wnt-3a, *vestigial tail* (*vt*), cause caudal truncations of the body axis^{5–7}. We observed a similar axial truncation in LRP6^{−/−} neonates in which vertebrae caudal to the lumbar regions are absent (Fig. 1b). This was first evident as a reduction in the size of the tailbud at 8.5 days post coitum (d.p.c.), with the loss of paraxial mesoderm and caudal somites at later stages (Fig. 2a, b). Sections through the tailbud of LRP6^{−/−} embryos revealed an excess of neural tissue and a corresponding loss of paraxial mesoderm (Fig. 2c, d), similar to what was observed in the targeted Wnt-3a mutant⁶. As in *vt*

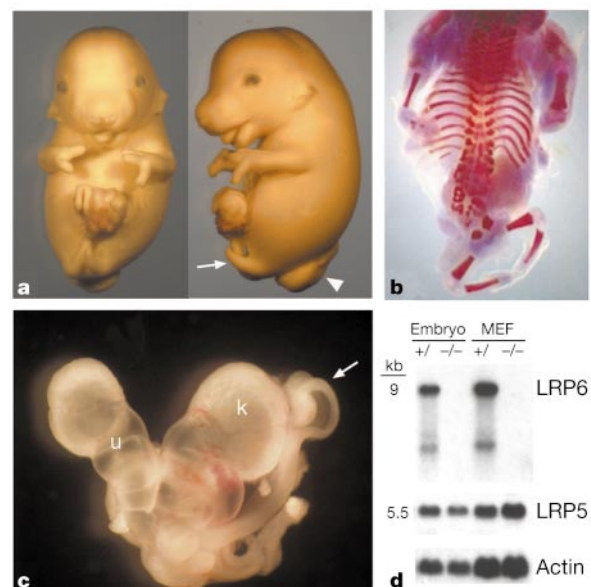


Figure 1 LRP6 mutant embryos exhibit severe developmental defects **a**, Embryo at 15.5 d.p.c. showing spina bifida (arrowhead), absence of a tail, malformed fore and hindlimbs (arrow) and retinal coloboma. **b**, Skeleton of a neonate showing truncation of the axial skeleton and loss of distal limb structures. **c**, Urogenital system of an 18.5-d.p.c. female embryo with small, cystic kidneys (k), expanded ureter (u) and developmentally delayed ovaries (arrow). **d**, Northern blot of total RNA from non-mutant (+/) and mutant (−/−) 12.5-d.p.c. embryos and embryonic fibroblasts (MEF).

* Present addresses: Department of Molecular & Cell Biology, Harvard University, 16 Divinity Avenue, Cambridge, Massachusetts 02138, USA (J.B.); Department of Biochemistry, University of Leicester, Leicester LE1 7RH, UK (S.M.).

mutants⁷, Wnt-3a transcripts were reduced in the tailbud (Fig. 2e). By 10.5 d.p.c., Wnt-3a transcripts were completely absent from the tail (Fig. 2f), probably reflecting the depletion of mesodermal precursors at this stage. The extent of axial truncations in *vt* and Wnt-3a null mutants is sensitive to the dose of Wnt-3a (ref. 7). On the basis of the level at which the axial skeleton is truncated, we conclude that LRP6^{-/-} embryos display a severely hypomorphic Wnt-3a phenotype.

To determine whether the *vestigial tail* phenotype is modified by the dose of LRP6, we crossed *vt*;LRP6 double heterozygous animals to *vt/vt* animals and collected embryos at 11.5 (data not shown) and 13.5 d.p.c. (Fig. 3). We observed the expected ratio of phenotypic classes: half (23) showed a normal tail length (corresponding to the *vt/+*;+/+ and *vt/+*;LRP6^{+/+} genotypes) and half (21) exhibited a truncation of the tail (representing the *vt/vt*;+/+ and *vt/vt*;LRP6^{+/+} genotypes). The tail defects in *vt/vt* embryos were clearly more pronounced among all eight embryos that inherited the LRP6 mutant allele (compare Fig. 3c and d). The enhancement of the *vestigial tail* phenotype upon elimination of one functional copy of the LRP6 gene provides genetic evidence that LRP6 and Wnt-3a function in the same pathway.

We next examined mutant embryos for mid/hindbrain defects, a hallmark of Wnt-1-deficient mice^{8,9}. The phenotypes of a classical Wnt-1 allele, *swaying*, and a targeted mutation of Wnt-1 have variable expressivity¹⁰, exhibiting a range of malformations of the midbrain and cerebellum. About one-half of the LRP6 mutants exhibited neural tube closure defects (exencephaly and/or spina bifida), and among those embryos that were not exencephalic morphological changes at the mid/hindbrain junction were observed. At 10.5 d.p.c., the mid/hindbrain boundary was less distinct in LRP6 mutant embryos (Fig. 4a, b), and by 14.5 d.p.c., a deletion of the caudal midbrain was evident (Fig. 4c, d). At birth, the inferior colliculus was absent and the cerebellum was disorganized

(Fig. 4e, f). Furthermore, an elevation and expansion of Wnt-1 transcripts observed in *swaying* embryos¹¹ was also seen in LRP6^{-/-} embryos (Fig. 4g, h). Thus, LRP6-deficient mice phenocopy the less severe class of mid/hindbrain defects observed in Wnt-1 mutant mice.

LRP6 mutant embryos displayed both anteroposterior and dorsoventral patterning defects in the limbs. These defects are reminiscent of those observed in the Wnt-7a mutant mice^{12,13} which fail to maintain Shh expression in the zone of polarizing activity (ZPA) leading to the variable loss of posterior digits. Wnt-7a is also required to establish dorsal cell fates in the limb: in its absence, a biventral limb is formed. Because hindlimb defects in the LRP6 mutants may be exacerbated by a general disruption of caudal development, we focused our analysis on forelimb development. The forelimbs of LRP6^{-/-} embryos showed a loss of one or more posterior digits (Fig. 5a) that was presaged at earlier stages by a gradual loss of Shh expression (Fig. 5b). Moreover, some of the remaining digits contained ectopic tendons diagnostic of a ventralized limb (Fig. 5c, d). A distinguishing feature of the LRP6 limb phenotype not observed in Wnt-7a mutants was a failure to maintain the apical ectodermal ridge (AER), as judged by discontinuous expression of Fgf8 (Fig. 5e, f). Discontinuity of the AER may explain the variable loss of anterior digits observed in LRP6 mutant embryos. The added complexity may indicate that, in addition to Wnt-7a, LRP6 mediates signalling by other Wnts in the limb.

Clearly, LRP6 mutant mice do not exhibit all of the Wnt phenotypes reported to date (for example, the early embryonic lethality associated with Wnt-3 mutants¹⁴ and the limb outgrowth defects in Wnt-5a mutants¹⁵); therefore, LRP6 may be involved in signalling by only a subset of vertebrate Wnts. Furthermore, the defects in LRP6 mutant embryos are generally less severe than those exhibited by individual Wnt mutants. Notably, a second highly

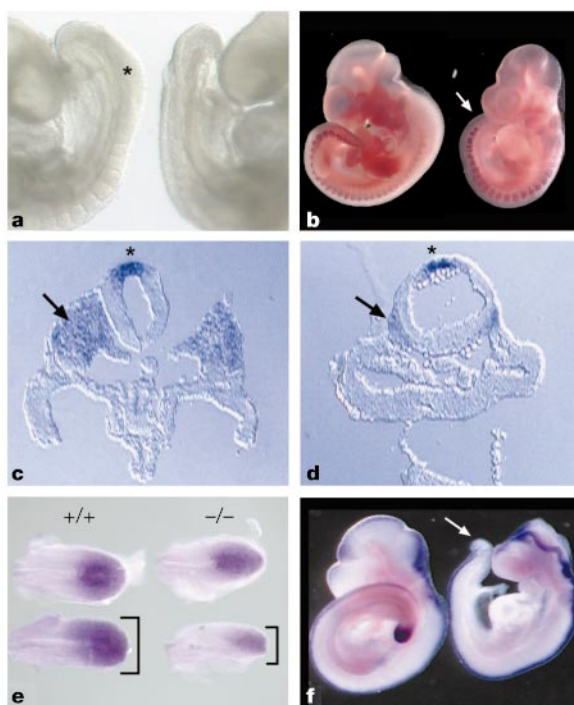


Figure 2 Abnormal tailbud and caudal somite formation in LRP6-deficient embryos. **a**, Embryos at 9.5 d.p.c. showing reduced paraxial mesoderm (asterisk) in the mutant (right). **b**, Myf5 expression in 10.5-d.p.c. embryos showing the loss of caudal somites (arrow) in the mutant (right). **c**, **d**, Cross-sections of 9.5-d.p.c. wild-type (**c**) and mutant (**d**) tailbuds stained for Pax3 expression. Mutants show an expansion of the neural tube

(asterisk) and a reduction of presomitic mesoderm (arrow). **e**, Mutant ($-/-$) tailbuds at 8.5 d.p.c. are narrower with reduced expression of Wnt-3a. **f**, Wnt-3a transcripts are undetectable in the tail (arrow) of 10.5 d.p.c. mutant embryos (right), but maintained in dorsal neural tube.

related family member, LRP5 (ref. 16), is widely expressed in embryos (Fig. 1c; and data not shown) and may limit the range and severity of Wnt phenotypes observed in $LRP6^{-/-}$ mice.

Our data suggest that LRP6 is required for efficient signalling by several Wnt proteins. This hypothesis is supported by a study of the *arrow* mutation in *Drosophila*¹⁷ and by overexpression of LRP6 in *Xenopus* embryos¹⁸, which corroborate an evolutionary conserved role for LRP6 in Wg/Wnt signal transduction. Members of the LDL receptor family are classically known for their involvement in the endocytic pathway, mediating the uptake through clathrin-coated pits of macromolecules such as apolipoproteins, lipases and proteases^{19,20}. Mice carrying mutations in two members of the LDL receptor family (the VLDL and ApoER2 receptors) have been shown to phenocopy mouse mutations in *reelin* or *disabled*, providing the first indication that LDL receptors can participate in specific signalling pathways²¹. Given the central importance of the LDL receptor family in the transport of lipoproteins and other macromolecules into the cell, the relationship between nutrient uptake and the response of cells to developmental signals such as Wnts warrants further investigation. □

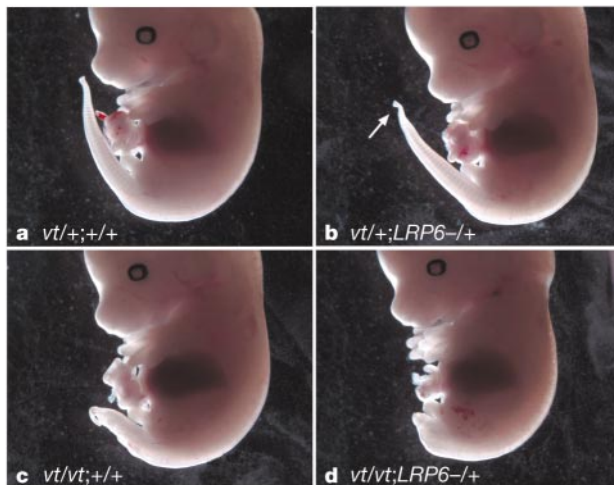


Figure 3 The LRP6 mutation enhances the *vestigial tail* phenotype. **a–d**, Embryos at 13.5 d.p.c. collected from matings between *vt/vt* and *vt/+;LRP6^{-/+}* animals. The genotypes of the embryos are given. The penetrance of tail kinks, seen occasionally in both *vt* and LRP6 heterozygotes, is increased in double heterozygotes (arrow in **b**).

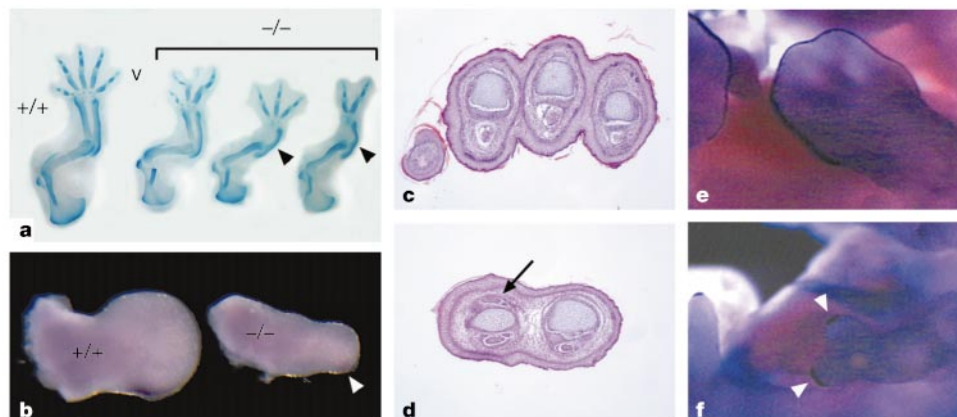


Figure 5 LRP6 mutant limbs show patterning and AER defects. **a**, Forelimb skeletons of 15.5 d.p.c. wild-type (+/+) and mutant (-/-) embryos. Mutant limbs show a consistent loss of the most posterior digit (V). Deletion of additional digits and the radius, as well as malformation of the ulna (arrowheads) is variable. **b**, Shh expression in wildtype (+/+) and mutant (-/-) 10.5-d.p.c. forelimb buds. Arrowhead indicates residual Shh transcripts in

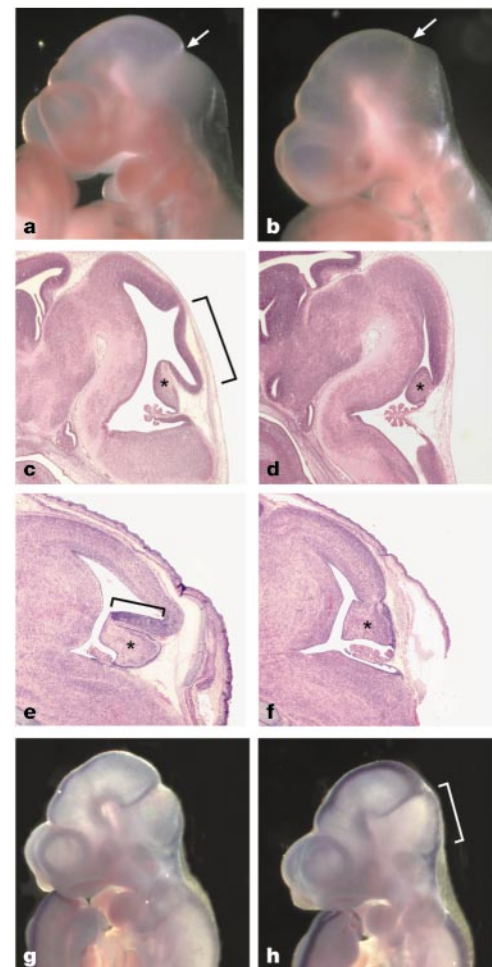


Figure 4 Caudal midbrain and cerebellar defects in LRP6 mutant embryos. **a, b**, The isthmus (arrow) demarcating the mid/hindbrain junction of wild-type embryos (**a**) is less distinct in mutants (**b**) at 10 s.d.p.c. **c, d**, Sagittal sections through the brain of 14.5-d.p.c. wild-type (**c**) and mutant (**d**) embryos reveal a deletion of the caudal midbrain (bracket in **c**) and cerebellar dysmorphology (asterisk) in the mutant. **e, f**, Similar sections of wild-type (**e**) and mutant (**f**) neonates show a loss of the inferior colliculus (bracket in **e**) and malformation of the cerebellum (asterisk) in the mutant. **g, h**, At 10.5 d.p.c., Wnt-1 transcript levels are increased overall and expanded into the metencephalon of the mutant (bracket in **h**) compared with wild type (**g**).

the mutant. **c, d**, Cross sections of 17.5-d.p.c. wild-type (**c**) and mutant (**d**) forelimb digits show a dorsal duplication of the ventral tendon in the mutant (arrow). **e, f**, Compared with wild type (**e**), Fgf8 expression is discontinuous in the AER of 12.5-d.p.c. mutant forelimbs (arrowheads in **f**).

Methods

Generation and genotyping of LRP6 mutant mice

Mouse embryonic stem cells (CGR8.8, derived from the 129/Ola mouse strain) were electroporated with the secretory trap vector pGT1.8TM as described⁴. We identified a cell line (Ex187) carrying an insertion in LRP6 by direct sequencing of complementary DNAs obtained by 5' rapid amplification of cDNA ends (RACE)²². Germline chimaeras generated by blastocyst injection were test bred and backcrossed to C57/BL6 mice for four generations before setting up intercrosses of heterozygous mutant animals. To genotype animals at weaning, dot blots of DNA prepared from tail biopsies were probed with vector (β geo) sequences²³. Mutant embryos before 8.5 d.p.c. were genotyped by the X-gal staining⁴ intensity of yolk sacs. Mutant embryos at later stages were identified by morphology or X-gal staining. Murine embryonic fibroblasts were isolated from eviscerated 12.5- and 14.5-d.p.c. embryos using standard procedures²⁴ and were genotyped by X-gal staining intensity. Northern blot analysis of midgestation embryos and fibroblasts was used to provide independent confirmation of genotypes scored by X-gal staining. Probes: LRP6, a 1.9-kbp BglII cDNA fragment (3' of the insertion); LRP5, a 400-bp EcoRI/NorI cDNA fragment containing the 3' untranslated region (mouse expressed sequence tag, Genebank no. A1119858); β -actin, 610-bp PstI cDNA fragment.

Skeletal preparation and general histology

We carried out skeletal preparations by standard methods¹². For histology of tissue sections, embryos were fixed in Bouin's solution for 5–12 h (depending on size of embryo), dehydrated and embedded in paraffin, and 8- μ m sections were stained with haematoxylin and eosin according to standard protocols.

vt/LRP6 matings

vt/vt mice (Jackson Laboratories, stock no. 000553) were crossed with LRP6 heterozygotes to produce vt/+;LRP6⁺/+ and vt/+;+/+ offspring. Compound heterozygotes were subsequently mated to vt homozygotes. Inheritance of the LRP6 mutant allele was determined by β -gal staining of yolk sacs or dot blots of tail DNA as described above.

In situ hybridization

Whole-mount *in situ* hybridization of mouse embryos was carried out essentially as described²⁵. Whole-mount embryos were sectioned on a cryostat. Probes were provided by A. McMahon (Pax3, Wnt-3a, Wnt-1 and Shh), G. Martin (Fgf8) and R. Harland (Myf-5).

Received 21 June; accepted 29 August 2000.

- Wodarz, A. & Nusse, R. Mechanisms of Wnt signalling in development. *Annu. Rev. Cell Dev. Biol.* **14**, 59–89 (1998).
- Uusitalo, M., Heikkilä, M. & Vainio, S. Molecular genetic studies of Wnt signalling in the mouse. *Exp. Cell Res.* **253**, 336–348 (1999).
- Brown, S. D. *et al.* Isolation and characterization of LRP6, a novel member of the low density lipoprotein receptor gene family. *Biochem. Biophys. Res. Commun.* **248**, 879–888 (1998).
- Skarnes, W. C., Moss, J. E., Hurlley, S. M. & Beddington, R. S. P. Capturing genes encoding membrane and secreted proteins important for mouse development. *Proc. Natl Acad. Sci. USA* **92**, 6592–6596 (1995).
- Takada, S. *et al.* Wnt-3a regulates somite and tailbud formation in the mouse embryo. *Genes Dev.* **8**, 174–189 (1994).
- Yoshikawa, Y., Fujimori, T., McMahon, A. P. & Takada, S. Evidence that absence of Wnt-3a signalling promotes neuralization instead of paraxial mesoderm development in the mouse. *Dev. Biol.* **183**, 234–242 (1997).
- Greco, T. L. *et al.* Analysis of the *vestigial tail* mutation demonstrates that Wnt-3a gene dosage regulates mouse axial development. *Genes Dev.* **10**, 313–324 (1996).
- McMahon, A. P. & Bradley, A. The Wnt-1 (int-1) proto-oncogene is required for development of a large region of the mouse brain. *Cell* **62**, 1073–1085 (1990).
- Thomas, K. R. & Capocchi, M. R. Targeted disruption of the murine int-1 proto-oncogene resulting in severe abnormalities in midbrain and cerebellar development. *Nature* **346**, 847–850 (1990).
- Thomas, K. R., Musci, T. S., Neumann, P. E. & Capocchi, M. R. *Swaying* is a mutant allele of the proto-oncogene Wnt-1. *Cell* **67**, 969–976 (1991).
- Bally-Cuif, L., Chollet, B. & Wassef, M. Involvement of Wnt-1 in the formation of the mes/metencephalic boundary. *Mech. Dev.* **53**, 23–34 (1995).
- Parr, B. A. & McMahon, A. P. Dorsalizing signal Wnt-7a required for normal polarity of D–V and A–P axes of mouse limb. *Nature* **374**, 350–353 (1995).
- Yang, Y. & Niswander, L. Interaction between the signalling molecules WNT7a and SHH during vertebrate limb development: dorsal signals regulate anteroposterior patterning. *Cell* **80**, 939–947.
- Liu, P. *et al.* Requirement for Wnt3 in vertebrate axis formation. *Nature Genet.* **22**, 361–365 (1999).
- Yamaguchi, T. P., Bradley, A., McMahon, A. P. & Jones, S. A. Wnt5a pathway underlies outgrowth of multiple structures in the vertebrate embryo. *Development* **126**, 1211–1223 (1999).
- Hey, P. J. *et al.* Cloning of a novel member of the low-density lipoprotein receptor family. *Gene* **216**, 103–111 (1998).
- Wherli, M. *et al.* *arrow* encodes an LDL-receptor-related protein essential for Wingless signalling. *Nature* **407**, 527–530 (2000).
- Tamai, K. *et al.* LDL receptor-related proteins in Wnt signal transduction. *Nature* (this issue).
- Brown, M. S., Herz, J. & Goldstein, J. L. Calcium cages, acid baths and recycling receptors. *Nature* **388**, 629–630 (1997).
- Willnow, T. E. The low-density lipoprotein receptor gene family: multiple roles in lipid metabolism. *J. Mol. Med.* **77**, 306–315 (1999).
- Trommsdorff, M. *et al.* Reeler/Disabled-like disruption of neuronal migration in knockout mice lacking the VLDL receptor and ApoE receptor 2. *Cell* **97**, 689–701 (1999).

- Townley, D. J., Avery, B. J., Rosen, B. & Skarnes, W. C. Rapid sequence analysis of gene trap integrations to generate a resource of insertional mutations in mice. *Genome Res.* **7**, 293–298 (1997).
- Brennan, J. & Skarnes, W. C. in *Methods in Molecular Biology, Molecular Embryology: Methods and Protocols* Vol. 97 (eds Sharpe, P. & Mason, I.) 123–138 (Humana Press, Totowa, NJ, 1999).
- Hogan, B., Beddington, R., Constantini, F. & Lacy, E. (eds) *Manipulating the Mouse Embryo: A Laboratory Manual* 2nd edn 260–261 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1994).
- Wilkinson, D. G. & Nieto, M. A. Detection of messenger RNA by *in situ* hybridization to tissue sections and whole mounts. *Methods Enzymol.* **225**, 361–367 (1993).

Acknowledgements

We wish to thank F. Hess for providing us with the mouse LRP5 and LRP6 cDNAs; V. Wilson, J. Rubenstein and T. Greco for helpful discussions; A. Peterson, J. Rine, G. Garriga and members of the lab for critical comments on the manuscript. This work was funded in part by the Biotechnology and Biological Sciences Research Council (UK), the Chicago Community Trust and a grant from the March of Dimes. W.C.S. is a 1998 Searle Scholar.

Correspondence and requests for materials should be addressed to W.C.S. (e-mail: skarnes@socrates.berkeley.edu).

Interaction of oestrogen receptor with the regulatory subunit of phosphatidylinositol-3-OH kinase

Tommaso Simoncini*, Ali Hafezi-Moghadam†, Derek P. Brazil‡, Klaus Ley†, William W. Chin§ & James K. Liao*

* Cardiovascular and § Genetics Divisions, Department of Medicine, Brigham & Women's Hospital and Harvard Medical School, 221 Longwood Avenue, Boston, Massachusetts 02115, USA
† Department of Biomedical Engineering, University of Virginia Health Sciences Center, Charlottesville, Virginia 22908, USA
‡ Howard Hughes Medical Institute, Joslin Diabetes Center, Boston, Massachusetts 02215, USA

Oestrogen produces diverse biological effects through binding to the oestrogen receptor (ER)¹. The ER is a steroid hormone nuclear receptor, which, when bound to oestrogen, modulates the transcriptional activity of target genes². Controversy exists, however, concerning whether ER has a role outside the nucleus³, particularly in mediating the cardiovascular protective effects of oestrogen⁴. Here we show that the ER isoform, ER α , binds in a ligand-dependent manner to the p85 α regulatory subunit of phosphatidylinositol-3-OH kinase (PI(3)K). Stimulation with oestrogen increases ER α -associated PI(3)K activity, leading to the activation of protein kinase B/Akt and endothelial nitric oxide synthase (eNOS). Recruitment and activation of PI(3)K by ligand-bound ER α are independent of gene transcription, do not involve phosphotyrosine adapter molecules or *src*-homology domains of p85 α , and extend to other steroid hormone receptors. Mice treated with oestrogen show increased eNOS activity and decreased vascular leukocyte accumulation after ischaemia and reperfusion injury. This vascular protective effect of oestrogen was abolished in the presence of PI(3)K or eNOS inhibitors. Our findings define a physiologically important non-nuclear oestrogen-signalling pathway involving the direct interaction of ER α with PI(3)K.

PI(3)K mediates the cellular effects of platelet-derived growth factor (PDGF)⁵, insulin⁶ and vascular endothelial growth factor (VEGF)⁷. The predominant form of PI(3)K comprises p85 α , an adapter/regulatory subunit of relative molecular mass 85,000 (M_r 85K), and p110, a catalytic subunit⁸ of M_r 110K. PI(3)K catalyses the

formation of lipid mediators^{9,10} which recruit signalling molecules containing phosphatidylinositol (PtdIns)-3,4,5-P₃-binding or pleckstrin homology domains such as phosphatidylinositol-dependent kinases and protein kinase Akt^{11,12}. The activation of Akt through phosphorylation of Thr 308/Ser 473 (ref. 13) mediates many of the downstream cellular effects of PI(3)K, including stimulation of glucose transporter-4 membrane translocation¹⁴, inactivation of glycogen synthase kinase-3 (ref. 15), and activation of eNOS^{16,17} and cell survival pathways¹⁸. Although oestrogen stimulates eNOS activity¹⁹ and promotes cell survival, it is not known whether PI(3)K mediates these effects of oestrogen.

In human vascular endothelial cells, physiological concentrations of 17 β -oestradiol (E₂) increased eNOS activity in a biphasic manner (effector concentration for half-maximal response (EC₅₀) $\approx$ 0.1 nM) (Fig. 1a, b). The initial increase was mediated by mitogen-activated protein (MAP) kinases¹⁹; the second increase was completely blocked by the PI(3)K inhibitor, wortmannin. The increase in eNOS activity was also blocked by the ER antagonist ICI 182,780; and the inactive E₂ stereoisomer 17 α -oestradiol (α E₂) had no effect. In murine fibroblasts transfected with ER α and eNOS complementary DNAs, E₂ produced an eightfold increase in eNOS activity in wild-type but not in p85 α -deficient (p85 α ^{-/-}) fibroblasts²⁰ (Fig. 1c). Furthermore, in p85 α ^{-/-} fibroblasts co-transfection of p85 α cDNA led to a fourfold increase in E₂-stimulated eNOS

activity, whereas in wild-type fibroblasts co-transfection of a dominant-negative p85 α mutant cDNA decreased E₂-stimulated eNOS activity by more than 50%.

In non-transfected human endothelial cells, E₂ increased endogenous PtdIns-3,4,5-P₃ levels in a time-delayed manner similar to the wortmannin-sensitive phase of eNOS activation (Fig. 2a). In contrast, insulin rapidly increased endogenous PtdIns-3,4,5-P₃ levels⁶ and eNOS activity²¹. Increases in PtdIns-3,4,5-P₃ levels correlated temporally with the ligand-dependent increases in ER α -associated PI(3)K activity (Fig. 2b); events that were blocked by ICI 182,780 and wortmannin (Fig. 2c). Consistent with a rapid, non-nuclear effect of ER on eNOS activation, E₂-stimulated PI(3)K activity was blocked by another ER antagonist, tamoxifen, but not by the MAP kinase inhibitor PD 98059, or by the transcriptional inhibitor actinomycin D (Fig. 2d). Insulin, which uses the phosphotyrosine (p-Tyr) adapter molecule, insulin receptor substrate (IRS)-1, to interact with PI(3)K, increased PI(3)K activity in the p-Tyr and IRS-1 immunoprecipitate (Fig. 2e), but did not increase or augment E₂-associated PI(3)K activity. In contrast, E₂ did not increase p-Tyr- or IRS-1-associated PI(3)K activity (Fig. 2e). These findings suggest that ER α does not recruit PI(3)K that has been already activated by insulin, and that PI(3)K activation by ER and IRS-1 occurs through different mechanisms. Notably, the activation of PI(3)K extended to other steroid hormone nuclear receptors such as the thyroid hormone and glucocorticoid receptors (Fig. 2f). These interactions may explain some of the previously unrecognized functions of these nuclear hormone receptors.

ER α interacted with p85 α in a ligand-dependent manner in both non-transfected endothelial cells (Fig. 3a) and p85 α ^{-/-} fibroblasts

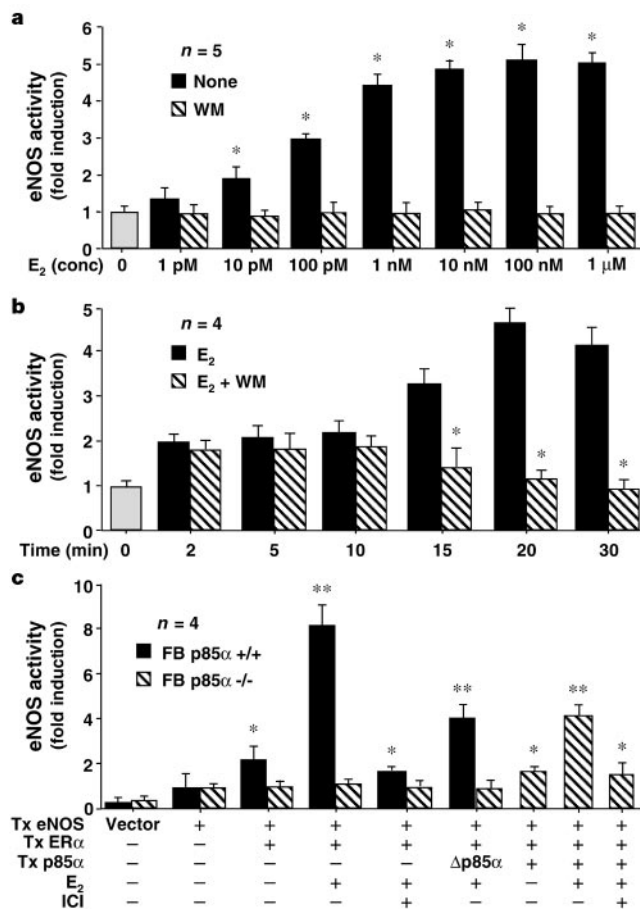


Figure 1 Activation of eNOS by oestrogen is mediated by PI(3)K. **a**, **b**, Concentration-dependent (**a**) and time-dependent (**b**) effects of E₂ and wortmannin (WM, 30 nM) on eNOS activity (fold induction versus baseline) in human vascular endothelial cells. Asterisk indicates $P < 0.05$ compared with unstimulated or E₂ stimulation. **c**, E₂-stimulated NOS activity in murine p85 α ^{+/+} and p85 α ^{-/-} fibroblasts (FB) transfected (Tx) with vector (pcDNA3), eNOS, ER α , p85 α or dominant-negative p85 α (Δ p85 α) cDNAs. Asterisk indicates $P < 0.05$ compared with transfection with eNOS cDNA alone; two asterisks indicate $P < 0.05$ compared with transfection with ER α and eNOS cDNAs.

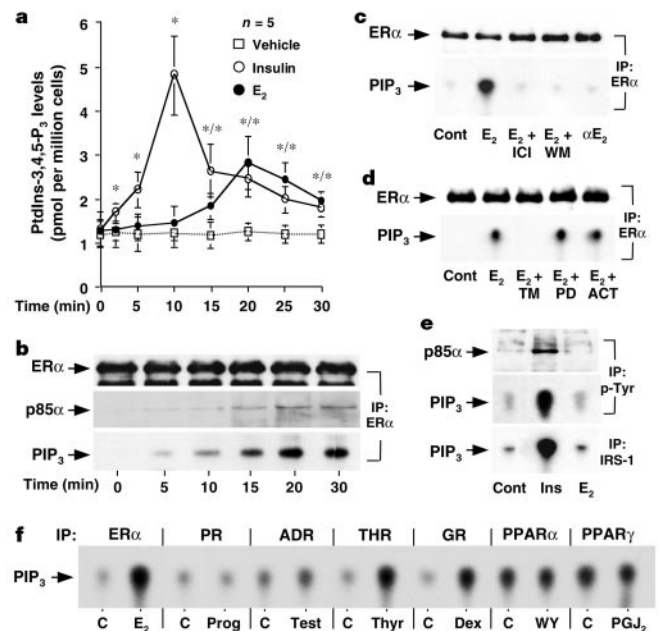


Figure 2 Oestrogen stimulates ER α -associated PI(3)K activity. **a**, Effect of vehicle (ethanol 0.01% v/v), E₂ (10 nM) or insulin (100 nM) on endogenous PtdIns-3,4,5-P₃ levels. Asterisk indicates $P < 0.05$ compared with vehicle. **b**, Time-dependent effect of E₂ on ER α , p85 α and PI(3)K activity (PIP₃) in ER α immunoprecipitate (IP). **c**, Effect of ICI (10 μ M) or WM on E₂ or 17 α -oestradiol (α E₂)-stimulated ER α -associated PI(3)K activity. Cells were pre-treated with ICI or WM for 30 min. **d**, Effect of tamoxifen (TM, 1 μ M), PD 98059 (PD, 5 μ M) and actinomycin D (ACT, 5 μ M) on ER α -associated PI(3)K activity. Inhibitors were added 2 h before E₂ stimulation. **e**, Effect of E₂ or insulin (Ins) on p-Tyr- and IRS-1-associated PI(3)K activity. **f**, Effect of E₂, progesterone (Prog, 10 nM), testosterone (Test, 10 nM), thyroid hormone (Thyr, 10 nM), dexamethasone (Dex, 1 μ M), WY14643 (WY, 100 μ M) and 15-deoxy- Δ ^{12,14}-prostaglandin J₂ (PGJ₂, 100 μ M) on PI(3)K activity in the corresponding steroid hormone nuclear receptor immunoprecipitates.

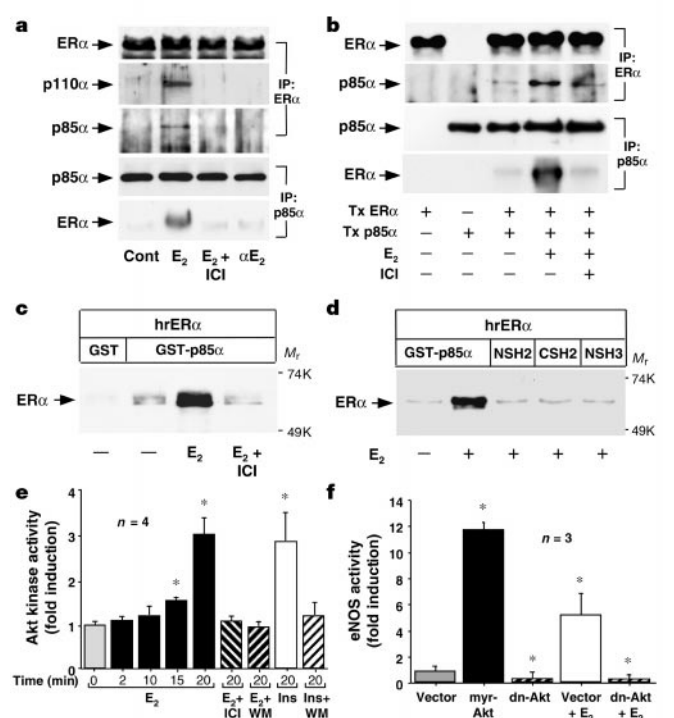


Figure 3 Ligand-dependent interaction of ER α with p85 α . **a**, **b**, Effect of E₂ on ER α –p85 α co-immunoprecipitation in non-transfected human endothelial cells (**a**) and murine p85 α ^{−/−} fibroblasts (**b**) transfected (Tx) with ER α and p85 α , alone or in combination. **c**, **d**, Affinity purification using agarose-conjugated GST or GST–p85 α (**c**), or GST–p85 α amino-terminal SH2 domain (NSH2, amino acids 321–470), carboxy-terminal SH2 domain (CSH2, 576–724), or SH3 domain (NSH3, 1–80) fusion protein and human recombinant (hr) ER α (**d**). **e**, E₂- or insulin (Ins)-stimulated Akt kinase activity. Asterisk indicates $P < 0.05$ compared with no stimulation. **f**, Effect of E₂ on eNOS activity (fold induction over baseline) in endothelial cells transfected with adenovirus containing no Akt (vector), constitutively active (myr), or a dominant-negative (dn) Akt. Asterisk indicates $P < 0.05$ compared with vector alone.

transfected with ER α and p85 α cDNAs (Fig. 3b). This ligand-dependent interaction was blocked by ICI 182,780 and was absent in p85 α ^{−/−} fibroblasts transfected with ER α cDNA alone. However, the ER isoform ER β , which is thought to mediate some of the cardiovascular effects of oestrogen⁴, did not interact with p85 α or recruit PI(3)K activity after E₂ stimulation (see Supplementary Information). The interaction of ER α and p85 α also occurred in the absence of adapter molecules or accessory proteins, as human recombinant ER α could still interact with glutathione S-transferase (GST)–p85 α fusion protein in a ligand-dependent manner in a cell-free system (Fig. 3c). This interaction, however, does not involve the src-homology SH2/SH3 domains of p85 α (Fig. 3d) which interact with p-Tyr residues of growth hormone receptors and adapter molecules^{22,23}. Heat shock protein 90, which binds and facilitates the function of ER²⁴ and eNOS²⁵, inhibited the interaction of ER α and p85 α .

The generation of PtdIns-3,4,5-P₃ leads to the recruitment and activation of Akt^{11,26}. E₂ stimulated Akt kinase activity in a time-delayed manner (Fig. 3e), similar to the increases observed in PtdIns-3,4,5-P₃ levels and eNOS activity. To determine whether E₂-stimulated eNOS activation is mediated by Akt, we transiently transfected bovine aortic endothelial cells with adenoviruses containing constitutively active (myr) and dominant-negative (dn) Akt mutants²⁷. Transfection of these cells with myr-Akt produced a substantial increase in eNOS activity, whereas overexpression of dn-Akt decreased basal eNOS activity below baseline and completely abolished E₂-stimulated eNOS activity (Fig. 3f).

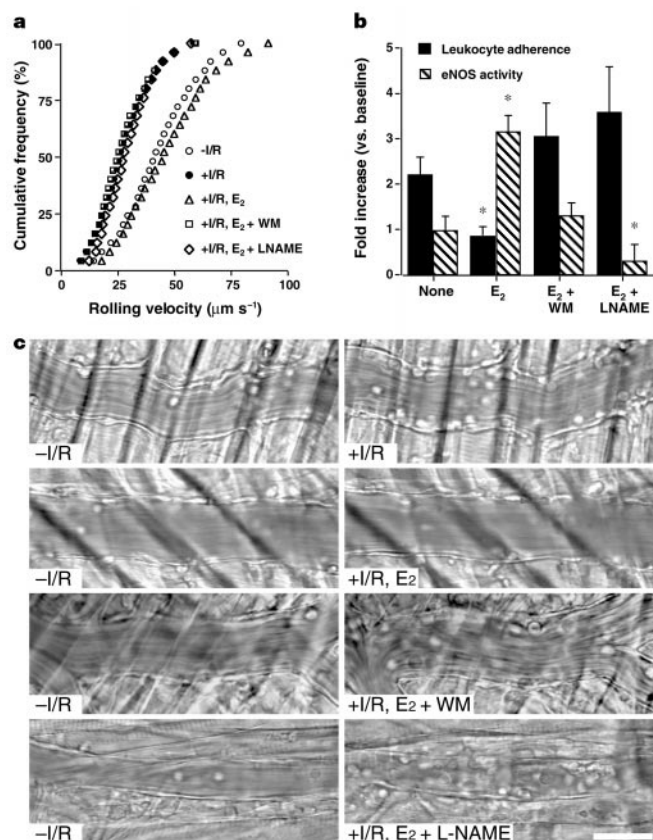


Figure 4 PI(3)K and NO mediate the vascular protective effects of oestrogen. Cumulative histograms of leukocyte rolling velocities before (−) and after (+) ischaemia and reperfusion (I/R) are shown. **a**, **b**, Effect of superfused WM (100 nM) or L-nitroarginine methylester (L-NAME, 0.1 mM) on leukocyte rolling velocity (**a**) and leukocyte adhesion and eNOS activity in the murine cremaster muscle (**b**). Data are expressed as fold increase over baseline before I/R in the same paired venules. Asterisk indicates $P < 0.001$ compared with untreated after I/R (None). **c**, Representative video images showing the same venules before (−) and after (+) I/R with the indicated treatments. Scale bar, 40 μ m.

To determine the physiological significance of this pathway, we used an established model of ischaemia and reperfusion (I/R) injury in the mouse cremaster muscle²⁸. I/R leads to leukocyte recruitment to the vascular wall, an event attenuated by NO and exacerbated by eNOS inhibitors such as L-nitroarginine methylester (L-NAME)²⁹. I/R reduced median leukocyte rolling velocity by 13.8 μ m s^{−1} ($P < 0.003$) and induced a 2.2-fold increase in the number of adherent leukocytes ($P < 0.001$) (Fig. 4a, b). Treatment with E₂ increased eNOS activity 3.2-fold and prevented the subsequent changes in leukocyte accumulation and rolling velocity after I/R. When wortmannin or L-NAME was applied to the cremaster muscle, measurements of leukocyte rolling velocity and accumulation were not different between untreated and E₂-treated mice after I/R, although L-NAME decreased eNOS activity below that of untreated mice (Fig. 4a–c). These findings indicate that the NO-induced vascular protective effect of oestrogen is predominantly mediated by PI(3)K.

Although the nuclear function of ER is clearly established, previous studies regarding the membrane and cytoplasmic effects of oestrogen remain inconclusive³. Linking the ER to PI(3)K suggests that the ER may be involved in a critical function outside the nucleus. In addition, the potential biological effects of oestrogen are considerably broadened because PI(3)K is known to mediate various cellular functions¹⁸. Although most of the ER is localized to the nucleus, we found that there is an increased level of membrane

and cytoplasmic ER after E₂ stimulation (data not shown). Indeed, a study has suggested that membrane-associated ER is involved in mediating NO release from endothelial cells³⁰. Thus, it is likely that PI(3)K is being recruited and activated by a small subset of ligand-bound, membrane-associated ERs. It remains to be determined, however, whether oestrogen can also activate PI(3)K indirectly, and whether PI(3)K can account for other rapid, non-nuclear effects of oestrogen. Further studies characterizing the interaction domains of ER α and p85 α should help clarify these issues. □

Methods

Cell cultures

Human and bovine aortic endothelial cells were obtained enzymatically with Type IA collagenase (1 mg ml⁻¹). They were cultured and stimulated under serum-starved conditions consisting of phenol-red-free Medium 199 (Gibco BRL, Life Technologies) with 0.4% charcoal-stripped fetal calf serum.

Immunoprecipitations

Cells were washed with ice-cold PBS and lysed with the following buffer: Tris-HCl (20 mM, pH 7.4), EDTA (10 mM), NaCl (100 mM), IGEAL (1%), Na₃VO₄ (1 mM), NaF (50 mM), PMSF (0.1 mg ml⁻¹) and aprotinin (0.3 mg ml⁻¹). We added the immunoprecipitating antibody (1 μ g) to equal amounts of cell lysates (0.5–1 mg) in 500 μ l of lysis buffer for 1 h at 4 °C with gentle rocking. Afterwards, 40 μ l of 1:1 Protein-A-agarose was added and the entire mixture was rocked gently for another 1 h at 4 °C. The mixture was then centrifuged at 12,000g for 5 min at 4 °C. The supernatant was removed and the immunoprecipitate was washed three times with 500 μ l of washing buffer, which differs from the lysis buffer in having 150 mM NaCl instead of 100 mM NaCl. We then separated proteins in the washed immunoprecipitate by SDS-PAGE and immunoblotted them with anti-ER α (Ab-10: Clone TE111.5D11, NeoMarkers, Fremont, CA) or anti-p85 α (Upstate Biotech., Lake Placid, NY) antibody.

GST fusion protein-affinity purification

Human recombinant GST–p85 α fusion protein or GST (Sigma) bound to glutathione-agarose beads (1 μ g protein per 20 μ l beads) was suspended in 400 μ l of *Escherichia coli* protein extract solution (10 mg ml⁻¹) and incubated with 1 μ g human recombinant ER α (Panvera, Madison, WI) for 1 h at 4 °C. We pelleted the samples, and washed the beads five times with a buffer containing 50 mM potassium phosphate, pH 7.5, 150 mM KCl, 1 mM MgCl₂, 10% (v/v) glycerol and 1% (v/v) Triton X-100 plus protease inhibitors. The beads were re-suspended in 50 μ l of 2 $\times$ Laemmli's buffer and boiled for 5 min. Proteins were separated on SDS-PAGE.

Model of vascular injury

Ten-week-old, 24 g, male C57BL/6 mice (Hilltop, Scottsdale, PA) were subcutaneously implanted with 1.5 mg of slow-release E₂ tablets (Innovative Research of America, Sarasota, FL) 3–5 days before experiments to ensure steady-state serum E₂ levels and to avoid any effects of surgery on baseline haemodynamic parameters. Mice implanted with E₂ tablets had a serum E₂ level of 760 $\pm$ 30 pg ml⁻¹ compared with that of vehicle-treated mice (24 $\pm$ 6 pg ml⁻¹). Mice were anaesthetized and the cremaster muscle was studied under intravital microscopy²⁸. Ischaemia was induced by applying pressure to supplying arteries just sufficient to stop blood flow for 30 min. In some experiments, wortmannin (100 nM) or L-NAME (0.1 mM) was applied to the cremaster muscle during the ischaemic period. The pressure was released for reperfusion, and the same vessels were recorded in each animal before and after I/R. The rolling velocities of 25 leukocytes were measured in each venule, sorted and averaged for each rank to construct cumulative histograms. The velocities of 3,750 leukocytes were measured in 150 venules before and after I/R. The number of firmly adherent leukocytes was measured before and after I/R in the same 200- μ m long segments of venules. The following number of venules were studied for leukocyte adhesion: untreated, 15 venules; E₂-treated, 20 venules; E₂-treated with wortmannin, 25 venules; E₂-treated with L-NAME, 15 venules. Cremaster eNOS activity was measured in three untreated, four E₂-treated, five E₂-treated with wortmannin and four E₂-treated with L-NAME mice.

Received 9 June; accepted 21 August 2000.

- Green, S. *et al.* Human oestrogen receptor cDNA: sequence, expression and homology to v-erb-A. *Nature* **320**, 134–139 (1986).
- Kumar, V. *et al.* Functional domains of the human estrogen receptor. *Cell* **51**, 941–951 (1987).
- Pietras, R. J. & Szego, C. M. Specific binding sites for oestrogen at the outer surfaces of isolated endometrial cells. *Nature* **265**, 69–72 (1977).

- Gustafsson, J. A. Novel aspects of estrogen action. *J. Soc. Gynecol. Investig.* **7**, S8–9 (2000).
- Escobedo, J. A. *et al.* cDNA cloning of a novel 85 kd protein that has SH2 domains and regulates binding of PI3-kinase to the PDGF beta-receptor. *Cell* **65**, 75–82 (1991).
- Ruderman, N. B., Kapeller, R., White, M. F. & Cantley, L. C. Activation of phosphatidylinositol 3-kinase by insulin. *Proc. Natl Acad. Sci. USA* **87**, 1411–1415 (1990).
- Papapetropoulos, A., Garcia-Cardena, G., Madri, J. A. & Sessa, W. C. Nitric oxide production contributes to the angiogenic properties of vascular endothelial growth factor in human endothelial cells. *J. Clin. Invest.* **100**, 3131–3139 (1997).
- Carpenter, C. L. *et al.* Purification and characterization of phosphoinositide 3-kinase from rat liver. *J. Biol. Chem.* **265**, 19704–19711 (1990).
- Auger, K. R., Serunian, L. A., Soltoff, S. P., Libby, P. & Cantley, L. C. PDGF-dependent tyrosine phosphorylation stimulates production of novel polyphosphoinositides in intact cells. *Cell* **57**, 167–175 (1989).
- Rameh, L. E. & Cantley, L. C. The role of phosphoinositide 3-kinase lipid products in cell function. *J. Biol. Chem.* **274**, 8347–8350 (1999).
- Stephens, L. *et al.* Protein kinase B kinases that mediate phosphatidylinositol 3,4,5- trisphosphate-dependent activation of protein kinase B. *Science* **279**, 710–714 (1998).
- Delcommenne, M. *et al.* Phosphoinositide 3-OH kinase-dependent regulation of glycogen synthase kinase 3 and protein kinase B/AKT by the integrin-linked kinase. *Proc. Natl Acad. Sci. USA* **95**, 11211–11216 (1998).
- Alessi, D. R. *et al.* Mechanism of activation of protein kinase B by insulin and IGF-1. *EMBO J.* **15**, 6541–6551 (1996).
- Kohn, A. D., Summers, S. A., Birnbaum, M. J. & Roth, R. A. Expression of a constitutively active Akt Ser/Thr kinase in 3T3-L1 adipocytes stimulates glucose uptake and glucose transporter 4 translocation. *J. Biol. Chem.* **271**, 31372–31378 (1996).
- Cross, D. A., Alessi, D. R., Cohen, P., Andjelkovich, M. & Hemmings, B. A. Inhibition of glycogen synthase kinase-3 by insulin mediated by protein kinase B. *Nature* **378**, 785–789 (1995).
- Dimmeler, S. *et al.* Activation of nitric oxide synthase in endothelial cells by Akt-dependent phosphorylation. *Nature* **399**, 601–605 (1999).
- Fulton, D. *et al.* Regulation of endothelium-derived nitric oxide production by the protein kinase Akt. *Nature* **399**, 597–601 (1999).
- Franke, T. F., Kaplan, D. R. & Cantley, L. C. PI(3)K: downstream AKTion blocks apoptosis. *Cell* **88**, 435–437 (1997).
- Chen, Z. *et al.* Estrogen receptor alpha mediates the nongenomic activation of endothelial nitric oxide synthase by estrogen. *J. Clin. Invest.* **103**, 401–406 (1999).
- Fruman, D. A. *et al.* Impaired B cell development and proliferation in absence of phosphoinositide 3-kinase p85 α . *Science* **283**, 393–397 (1999).
- Zeng, G. & Quon, M. J. Insulin-stimulated production of nitric oxide is inhibited by wortmannin. Direct measurement in vascular endothelial cells. *J. Clin. Invest.* **98**, 894–898 (1996).
- Fantl, W. J. *et al.* Distinct phosphotyrosines on a growth factor receptor bind to specific molecules that mediate different signaling pathways. *Cell* **69**, 413–423 (1992).
- Kapeller, R., Toker, A., Cantley, L. C. & Carpenter, C. L. Phosphoinositide 3-kinase binds constitutively to α / β -tubulin and binds to gamma-tubulin in response to insulin. *J. Biol. Chem.* **270**, 25985–25991 (1995).
- Picard, D. *et al.* Reduced levels of hsp90 compromise steroid receptor action *in vivo*. *Nature* **348**, 166–168 (1990).
- Garcia-Cardena, G. *et al.* Dynamic activation of endothelial nitric oxide synthase by Hsp90. *Nature* **392**, 821–824 (1998).
- Burgering, B. M. & Coffey, P. J. Protein kinase B (c-Akt) in phosphatidylinositol 3-OH kinase signal transduction. *Nature* **376**, 599–602 (1995).
- Fujio, Y. & Walsh, K. Akt mediates cytoprotection of endothelial cells by vascular endothelial growth factor in an anchorage-dependent manner. *J. Biol. Chem.* **274**, 16349–16354 (1999).
- Kanwar, S., Smith, C. W. & Kubes, P. An absolute requirement for P-selectin in ischemia/reperfusion-induced leukocyte recruitment in cremaster muscle. *Microcirculation* **5**, 281–287 (1998).
- Kubes, P., Suzuki, M. & Granger, D. N. Nitric oxide: an endogenous modulator of leukocyte adhesion. *Proc. Natl Acad. Sci. USA* **88**, 4651–4655 (1991).
- Stefano, G. B. *et al.* Cell-surface estrogen receptors mediate calcium-dependent nitric oxide release in human endothelia. *Circulation* **101**, 1594–1597 (2000).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank T. Uchida, A. J. Prorock and K. L. Thomas for technical assistance; M. White for providing IRS-1/2 antibodies; M. Brown for ER α antibody; D. Fruman and L. Cantley for murine p85 α fibroblasts, GST–p85 α and sub-domains; M. Kasuga for wild-type and dominant-negative p85 α cDNAs; and K. Walsh for adenovirus Akt mutants. This work was supported by grants from the National Institutes of Health, the Mary Horrigan Connors Center for Women's Health, the American Heart Association and the Scuola Superiore di Studi e di Perfezionamento "S. Anna".

Correspondence and requests for materials should be addressed to J.K.L. (e-mail: jliao@rics.bwh.harvard.edu).

Genes, chains and rats

Gene expression, quicker PCR and rodent observation in this week's haul.

Proxy-2000

Columbus Instruments www.colinst.com

Closely observed rats and mice

This newly developed capacitive proximity sensor can monitor minuscule lab animal movements, including respiration. The sensor has two inputs, one of low sensitivity, which provides a signal in the range 0–1 V, proportional to the size of the animal movement. The other, with high sensitivity, can monitor movements associated with respiration. Proxy-2000 requires two electrodes, one on the top of the animal cage, and one on the bottom. By placing the top electrode closer to the animal, sensitivity is increased. The proximity sensor was developed for studying apnoea in genetically defective animals, but can also be used to monitor activity during oxygen-consumption measurements, as well as other applications. Signals from the proximity sensor can be fed to the computer via A/D for further processing or storage. Columbus Instruments can also provide a voltage-to-frequency converter/counter, which facilitates integration of animal activity over time.

Reader Service No. 100

PATIPC database

STN International www.fiz-karlsruhe.de

An electronic tool for patent searchers and patent attorneys

The PATIPC database is the online version of the International Patent Classification (IPC) designed to classify worldwide property rights by subject areas. IPC is published by the World Intellectual Property Organization (WIPO) in Geneva and the German Patent and Trademark Office in Munich. The main purpose of the classification is to facilitate the search for technical facts. The database is offered by FIZ Karlsruhe, operator of the European Service Center of STN International.

Reader Service No. 101

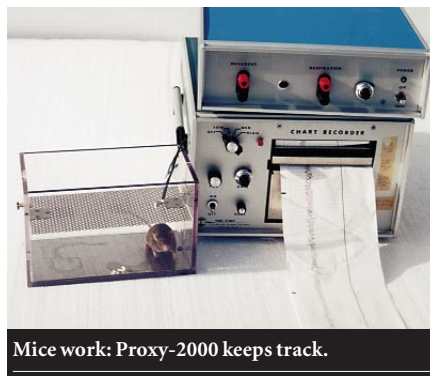
TotalLab

Phoretix www.totallab.com

How to get a press release noticed by science journals and lab equipment mags

The TotalLab program is an image analysis system for the analysis of 1D electrophoresis gels and 2D spots, dot and slow blots, other arrays and colonies. The demo software has recently been updated to allow potential users to evaluate the software using their own gels. And — here's the USP — for a limited period from 30 August every hundredth download will be given a free copy of the full TotalLab software.

Reader Service No. 102



Mice work: Proxy-2000 keeps track.

gNAPS

Whatman Bioscience www.whatman.com

Purification of genomic DNA from whole blood

gNAPS is based on a filter system that entraps the DNA in its matrix rather than binding it. This allows the separation process to be performed in a single, unidirectional reaction vessel, eliminating user interaction and multiple centrifugation steps. Samples require no pretreatment, and the entire process takes just 5 minutes, with a nominal 10-minute incubation stage that may be increased or decreased depending on the size of fragments required. Time and temperature of elution can be adjusted to produce varying proportions of high molecular weight DNA for a wide range of downstream applications, and it is not necessary to perform elution at elevated temperatures. And as there is no centrifugation, there should be no loss of white cells through premature rupture of the cell membranes.

Reader Service No. 103

BioMedicine

InforMax www.informaxinc.com

Bioinformatics software for drug discovery

Version 2.0 of SSBM (Software Solution For BioMedicine) includes new features and functionality for increasing productivity. The program is designed to perform high-speed analysis of both public and proprietary genetic databases within the security of the corporate firewall and provides speed, security and scalability through the employment of a client-server architecture including an Oracle relational database, UNIX Server and Java-based graphical user interface. This distributed architecture allows collaboration throughout a globally dispersed corporation (or network of several organizations) on multiple discrete research projects.

Reader Service No. 104

Matchmaker ESTs

Clontech www.clontech.com

Go to the library

A normalized human EST (expressed sequence tag) cDNA library, now available for GAL4-based two-hybrid analysis. The new Human Pretransformed Matchmaker EST cDNA Library combines the selection capabilities of Clontech's AH109 yeast strain with an EST library of over 240,000 ESTs, encompassing 26 human tissues in various developmental and disease states. Each sequence is expressed in all three reading frames, and is normalized so that rarely expressed genes are present at the same levels as those normally expressed at higher levels. By eliminating the amplification and library-scale transforma-

ADVERTISEMENTS

Custom Peptides & Polyclonal Antibodies

Please contact us about complete services for your custom peptide and polyclonal antibody needs.

Research Genetics

U.S. or Canada 800-533-4363

FAX 256-536-9016

Check out our homepage at
<http://www.resgen.com>

READER ENQUIRY NO. 39

Protein & Peptide Sequencing Service

by Biologica Co.

N-terminus Sequencing including In-gel digestion, Microfractionation
You only send us Gel spot or piece from 2D-E or PAGE containing a few pmol protein.



For Health and Science
Biologica co.

E-mail: biologic@rr.iij4u.or.jp
<http://www.iijnet.or.jp/biologica/>

READER ENQUIRY NO. 49

new on the market

tion steps (the EST library is pretransformed in yeast strain J693), the library screening process is reduced to only one or two weeks.
Reader Service No. 105

MasterAmp

Epicentre

www.epicentre.com

'Easy' but high-fidelity RT-PCR

Intended for users needing to clone and/or express an accurate RT-PCR product, the MasterAmp High Fidelity RT-PCR Kit is claimed to provide excellent accuracy with the option of a one-step protocol. The kit uses MMLV-RT Plus and also includes MasterAmp TAQurate DNA Polymerase, for a PCR step fidelity three times higher than Taq DNA polymerase. The dNTPs, MgCl₂ and buffer are all in one vial, and the kit contains both oligo (dT)₁₈ and random nonamer primers.

Reader Service No. 106

MultiBlock PCR

Hybaid

www.hybaidd.co.uk

New software for the MultiBlock PCR machine

New features included in Version 2.0 of Hybaid's Thermal Cycling Software include simultaneous downloading of programs, and the ability to run and pause up to 30 blocks with the new global control function. All GLP



Block vote: new software for MultiBlock.

and SOP requirements are now met with historical logging and real-time monitoring of any block. This includes a summary of a block's performance, and also the ability to review every cycle of any PCR run from any machine in absolute detail.

Reader Service No. 107

AdEasy

Quantum Appligene

www.quantum-appligene.com

Recombinant adenovirus production and expression

Combining the power of adenovirus technology with the simplicity and flexibility of *E. coli* genetics, the AdEasy system is a complete kit for recombinant adenovirus pro-

duction and gene expression. AdEasy extends the use of the adenovirus by eliminating the need for viral plaque formation. The adenovirus in this kit can accommodate 7.5 kb of foreign DNA, and is claimed to be capable of replication to the high titres useful for gene therapy applications.

Reader Service No. 108

FDD system

Hitachi

www.hitachi-soft.com

A complete system for FDD for gene expression detection

This FDD (fluorescence differential display) system from Hitachi Genetic Systems by MiraiBio, Inc. includes the FMBIOW II Fluorescence Imaging System, the RNAspectra Kit and the SA-43 Vertical Electrophoresis System. The makers claim a sensitivity equal to that of radioactive labelling. The new RNAspectra Kits feature a new generation mRNA Differential Display (DD) technology, developed in collaboration with GenHunter Corporation, using FDD instead of radioactivity. The kits, which come with the licence to practise the DD process, include everything necessary for FDD except Taq.

Reader Service No. 109

These notes are compiled in the Nature office from information provided by the manufacturers.

ADVERTISEMENTS

TREVIGEN

new

PARP

**Poly (ADP-ribose)
Polymerase Reagents**

- PARP inhibition assay
- Biotinylated NAD
- Recombinant human PARP
- Anti PARP and PAR

www.trevigen.com

US and Canada: 1 800 873 8443
TEL: 301 216 2800 • FAX: 301 216 2801

READER ENQUIRY NO. 60

**For Immunology Apoptosis,
& Cytokines Order Your Y2K
Immuno Kontakt Catalog**

UK Tel+44 1235 828200 Fax +44 1235 820482
D Tel+49 69 779099 Fax +49 69 13376880
CH Tel+41 91 6045522 Fax +41 91 6051785
amsbio@dial.pipex.com
www.immunok.com

READER ENQUIRY NO. 61

Database Management Systems

Data Update Services

Bioinformatics
Powered by GCG

- Comprehensive analysis packages
 - Enterprise-wide solutions
 - Extendable software
 - Computing security
 - Scriptable analysis

www.gcg.com

GCG is an Oxford Molecular Company

Client/Server & Desktop Software

Graphical Interfaces

READER ENQUIRY NO. 36

www.alphainnotech.com

**New Visions
in Digital Imaging**

Alpha Innotech Corporation

READER ENQUIRY NO. 48

TRANSFECTION REAGENTS

NEW

TRANSOME

SUV Liposome Reagent

GENETRANSFER

MLV Liposome Reagent

HMG-1,-2 MIXTURE
Efficient Gene Carrier

Wako

Wako USA Wako GmbH

877-714-1921 (49)02131 3110

www.wako-chem.co.jp

READER ENQUIRY NO. 37

Look no further

for the best value in

Oligo Synthesis

Competitive pricing

Routine 24 hour dispatch • Every Oligo QC verified by PAGE
Rigorous batch QC by PCR and Mass Spec
Web ordering and tracking • Local Freephone numbers
A great deal in any language!!

SIGMA GENOSYS (+44) (0) 1223 839000 • fax (+44) (0) 1223 839200
email: info@sigma-genosys.co.uk • www.sigma-genosys.co.uk

READER ENQUIRY NO. 33